ABSTRACT

Title of Dissertation:	A MULTI-OMICS APPROACH TO CHARACTERIZING THREE HEALTH RELEVANT FUNCTIONS OF THE HUMAN GUT MICROBIOME
	Domenick J. Braccia, Doctor of Philosophy, 2022
Dissertation directed by:	Dr. Brantley Hall, Department of Cell Biology and Molecular Genetics

The human gut is home to trillions of microorganisms that routinely interact with their human host in both beneficial and detrimental ways. The advent of next-generation sequencing and high-throughput "omics" technologies has created new opportunities to examine the role that the human gut microbiome plays on human health, especially in regard to gastrointestinal diseases such as Inflammatory Bowel Disease and colorectal cancer. In my dissertation, I utilize genomic, transcriptomic, metabolomic, and protein sequence datasets to characterize three health-relevant functions of the human gut microbiome.

First, I performed a multi-omic, bioinformatic analysis to identify the bacterial enzyme, bilirubin reductase. While bilirubin reduction to urobilinogen and stercobilinogen is a well-known function of the human gut microbiome, the enzyme(s) responsible for the conversion of bilirubin to non-toxic reduced products have yet to be fully characterized. In this chapter, I review how I leveraged publicly available metabolomic, metagenomic, and metatranscriptomic

data to explore over 2 million putative reductase genes and identify a candidate operon encoding bilirubin reductase.

Second, I examined sources of microbial hydrogen sulfide (H_2S) production by bacteria of the human gut microbiome. H_2S is a sulfuric gas produced by various bacterial phyla of the human gut microbiome and is implicated in the etiology of gastrointestinal diseases such as Inflammatory Bowel Disease and colorectal cancer. In this chapter, I show via bioinformatic analysis that the capacity to produce H_2S via cysteine degradation is ubiquitous in the human gut.

Third, I explored bacterial prodrug activation required for the activation of immune system modulators such as sulfasalazine. After curating amino acid sequences of known azoreducing genes and performing a protein sequence search across the Unified Human Gastrointestinal Genomes (UHGG) collection containing 4,644 genomes, I identified putative azoreducing and non-azoreducing bacterial strains to be experimentally validated. Together, these results highlight a successful mult-omic approach to characterizing three diverse but health-relevant functions of the human gut microbiome.

A MULTI-OMICS APPROACH TO CHARACTERIZING HEALTH RELEVANT FUNCTIONS OF THE HUMAN GUT MICROBIOME

by

Domenick J. Braccia

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor A. Brantley Hall, Chair
Professor Mihai Pop
Professor Rob Patro, Dean's Representative
Assistant Professor Philip Johnson
Associate Professor Stephen Mount

Dedication

This dissertation is dedicated to my late grandfather. He dedicated his life to making sure his grandchildren did not have to struggle the way he did. Rest in peace.

Acknowledgements

I want to acknowledge all of the friends, family, fellow students, staff and faculty that helped me get to getting to the end of my PhD here at the university of Maryland. This has been an extremely instrumental 5 years for my life and my career, and I owe a lot of thanks to a lot of mentors who helped me get to this place.

Thanks to Dr. Hector Corrada Bravo for mentoring me for the first three years of my PhD from 2017 to 2020. I learned so much from the lively and collaborative environment he built during those years, especially when it comes to coding in R, and best practices for biological data science and data visualization.

Thanks to Dr. Brantley Hall for mentoring me for the final two years of my PhD. I felt incredibly motivated by his lab's mission to modulate the human gut microbiome to positively affect human health, and I think each chapter of this dissertation reflects that. Also, thank you to Dr. Mihai Pop for co-advising me for these last two years. It was incredibly valuable to have insight from the computer science-focused end of computational biology.

Finally, I want to thank my amazing wife, Reena Philips. She never stopped believing in me, even through adversity. She is the rock of my life, and I know I would not be the same without her in my life.

Table of Contents

Dedication ii
Acknowledgements iii
Table of Contents iv
1. Introduction 1
1.1 Microbiome effect on human health 1
1.1.1 Overview 1
1.1.2 Processing of bile acids and pigments by human gut microbiota 2
1.1.3 Production of gases by human the gut microbiota 3
1.1.4 Modification of oral drugs by human gut microbiota 4
1.2 Contributions 6
2. Finding bilirubin reductase 8
2.1 Background 8
2.1.1 Constitutive bilirubin production and processing 8
2.1.2 Hyperbilirubinemia in neonates and adults 8
2.1.3 Potential therapies for neonatal jaundice 9
2.2 Methods and Results 10
2.2.1 Stratifying patient populations based on microbial bilirubin-reducing potential 10
2.2.2 Bacterial strain prioritization 12
2.2.3 Bilirubin reductase gene candidate prioritization 14
2.2.4 Experimental confirmation and comparative genomics analysis 14
2.2.5 Assessment of BRC4 abundance in BR and nonBR cohorts 15
2.3 Discussion 16
3. The capacity to produce hydrogen sulfide (H_2S) via cysteine degradation is ubiquitous in the human gut microbiome 20
3.1 Background 20
3.1.1 Endogenous H_2S production 23
3.2 Results 23
3.2.1 Cysteine-degrading genes are widely distributed in the human gut microbiome 23
3.2.2 Increased relative abundance of putative hydrogen sulfide-producing bacteria in the Inflammatory Bowel Disease and colorectal cancer gut microbiomes 24
3.2.3 Primary and secondary cysteine-degrading genes are actively expressed in the human gut microbiome 25
3.2.4 Core dissimilatory sulfate reduction genes and methanogenesis genes are co-expressed in vivo 26
3.3 Discussion 26
3.4 Methods 29
3.4.1 Curation of Profile Hidden Markov Models Involved in Cysteine-Degradation and Sulfate-Reduction 29

3.4.2 Search for Putative Hydrogen Sulfide Producing Bacteria in the Human Gut 30
3.4.3 Calculating relative abundances with Kraken 2 31
3.4.4 Transcriptomic Analysis of Hydrogen Sulfide Producing Genes and Methane Producing Genes 31
4. Gut microbiome-wide search for bacterial azoreductases reveals inter-individual differences in prodrug delivery potential to the intestines 38
4.1 Background 38
4.2 Results 41
4.2.1 Azoreductase genes are widely distributed in the human gut microbiome 41
4.2.2 Relative abundance of putative azoreducing bacteria is significantly higher than known azoreducing bacteria across healthy and disease populations 43
4.2.3 Relative abundance of putative azoreducing species fluctuates over time 44
4.3 Discussion 44
4.5 Methods 46
4.5.1 Curation of pHMMs representing azoreductase enzymes 46
4.5.2 Search for azoreductase genes across human gut microbial genomes 47
4.5.3 Relative abundance estimation of gut bacteria 47
4.5.4 Statistical analysis of relative abundance data from cMD, HMP2 and PRISM datasets 48
Figure 4.1. Mechanism of azoreduction by gut microbiota. 49
Figure 4.3. Azoreductases are widely distributed across gut bacterial taxonomy. . 51
Figure 4.4. Relative abundance of putative azoreducers exceeds that of known azoreducers across healthy, IBD, adenoma and CRC populations. 52
Figure 4.5. Known and putative azoreducing species are consistently more abundant in the IBD gut. 53
5. Conclusion 54
List of Figures 57
List of Abbreviations 58
Appendices 60
Bibliography 64

1. Introduction

The human body is home to trillions of microorganisms including bacteria (1), viruses (2), and some species of fungi (3) which interact with the host on a regular basis in both beneficial and detrimental ways (4). The skin, mouth, vagina, and gut are all home to thousands of different bacterial species which collectively are known as the human microbiome (5,6). Although the human microbiome encodes many fold more genes than the human genome itself, the vast majority of the genes encoded in the human microbiome have not been extensively characterized (7). Further, while many bacterial species and strains are shown to be statistically correlated with dozens of different health conditions including Inflammatory Bowel Disease (8), colorectal cancer (9), and even neurological disorders such as Alzheimer's disease (10,11), the mechanistic understanding of how and why certain microbes are associated with disease states is underdeveloped. By combining high-throughput sequencing technology and sequence processing algorithms with statistics and data science, real strides have been made in the discovery and development of novel microbial therapeutics for the treatment of gastrointestinal diseases such as Inflammatory Bowel Disease (12) and colorectal cancer (13).

1.1 Microbiome effect on human health

1.1.1 Overview

Over the last 20 years, progress has been made in elucidating the precise mechanisms by which hosts interact with their various microbiomes and vice versa. While most functions performed by the human gut microbiome have unknown effects on the host (14), there are certain functions that promote human health. For instance, gut microbes produce butyrate,

which is the preferred energy source of colonic epithelial cells (15). The gut microbiome also produces essential vitamins that are required for the proper functioning of certain proteins (16). Thirdly, interactions between commensal bacteria and the host both balance and educate the immune system. This is important so that the host immune system does not launch an immunological attack on benign inhabitants of the human gut (17,18). Finally, a healthy gut microbiome also confers colonization resistance, making it harder for pathogens and opportunistic microbes to gain a foothold in the gut (19). However, there are also functions of the human gut microbiome which deride human health. For instance, gut flora-dependent metabolism of choline to Trimethylamine (TMA) is linked to an increased incidence of heart disease (20). Some microbes can also modify drugs, such as digoxin, rendering them inactive (21). Gut bacteria also produce many volatile gases such as H_2S and CH_4 which are responsible for many common gastrointestinal symptoms such as bloating and sharp pains (22). Finally, dysbiosis of the gut microbiome can lead to imbalances in the immune system of genetically susceptible individuals, particularly in patients diagnosed with Inflammatory Bowel Disease (IBD) (23).

With both the beneficial and harmful effects of the human gut microbiome in mind, one can see that there is an enormous potential to promote human health through careful modulation of the gut microbiome.

1.1.2 Processing of bile acids and pigments by human gut microbiota

The human gut is a complex environment that receives input from various sources including the pancreas via the pancreatic duct, the gall bladder via the common bile duct, and components of the small intestine such as the duodenum, jejunum, and ileum (24). Bile acids

and pigments are dumped from the gall bladder into the duodenum via the common bile duct, where the bile acids take part in digestion of lipids and pigments such as bilirubin are readied for excretion via glucuronidation to conjugated bilirubin. After passing through the small intestine, conjugated bilirubin is deconjugated to bilirubin via beta-glucoronidases and is subsequently reduced to the urobilinoids urobilinogen and stercobilinogen by human gut microbiota (25). Disruption of one or more of these bilirubin processing pathways can lead to a condition known as hyperbilirubinemia, otherwise known as jaundice.

While bilirubin reduction to urobilinoids by the human gut microbiota has been known of for over 50 years (26), the gene or genes responsible for this process are yet to be identified and characterized. This is a surprising gap in scientific knowledge that I hope to address with the work done in this dissertation.

1.1.3 Production of gases by human the gut microbiota

Gas is a natural byproduct of bacterial fermentation of carbohydrates and amino acids with the most abundant gases being hydrogen (H_2), carbon dioxide (CO_2), and methane (CH_4) (27). While H_2 and CO_2 are produced by many gut-dwelling bacteria (28), CH_4 is produced by the genera *Methanobrevibacter* and *Methanosphaera*, with the most prominent species being *Mb. smithii* and *Msp. stadtmanae* (29). While odorous gases such as H_2S, indole, skatole, NH_3, and other volatile amines account for less than 1% of total flatus volume, their effect is quite noticeable. While all healthy individuals experience some level of gas production, excess gas production in the gut can lead to bloating, malodorous flatulence, and even degradation of the mucosal lining (30). Though many microbially produced metabolites have been measured and studied in the stool, there has been very little

research performed on gaseous metabolites present in flatus produced by the human gut microbiota. In chapter 3 of this dissertation, we explore the presence and relative abundance of H_2S -producing bacteria in the human gut and compare the capacity for H_2S production via two primary pathways: (i) cysteine degradation and (ii) dissimilatory sulfate reduction.

1.1.4 Modification of oral drugs by human gut microbiota

Although endogenous mechanisms of drug metabolism have been well established and studied, exogenous processing of drugs by the human microbiome has largely been overlooked. Orally administered drugs can undergo extensive modification by gut-dwelling microbes which can have a wide range of therapeutic outcomes. Generally, there are three types of direct mechanisms of microbial drug modification: (i) activation, (ii) inactivation, and (iii) toxification (31).

Examples of microbial activation of orally administered drugs include mesalamines such as sulfasalazine (32) and antibiotics such as prontosil (33). Both sulfasalazine and protonsil require the reduction of a nitrogen-nitrogen double to release the therapeutic moiety responsible for each drug's efficacy. For sulfasalazine, 5-aminosalicylic acid (5-ASA) is released and acts as an anti-inflammatory agent on inflamed tissues of individuals with Inflammatory Bowel Disease (34). For prontosil, triaminobenzene and sulfanilamide are released upon azo bond reduction and act together as a bacteriostatic antibiotic that inhibits folate metabolism (35,36). However, certain types of oral drugs can be inactivated by gut microbial metabolism

A prime example of microbial inactivation of drugs is digoxin. Digoxin is a cardiac glycoside traditionally prescribed for cardiac arrhythmia and heart failure (37). In about 10%

of patients, digoxin is metabolized to its clinically inactive form dihydrodigoxin resulting from bacterial reduction of the α,β-unsaturated lactone ring (21). Initially, the inactivation of methotrexate was thought to be caused by specific strains of *Eggerathella lenta*, but neither the presence nor abundance of *E. lenta* was shown to be predictive of digoxin inactivation (38–40). This is thought to be due to strain variants containing or not containing the membrane-bound electron shuttling operon, *cgr* (41).

Finally, gut microbiota are responsible for toxic effects of the anti-cancer drug CPT-11 and the NSAID acetaminophen. CPT-11 is an intravenous drug used in the treatment of colorectal cancer which is metabolized by host carboxylesterases to the bioactive SN-38, which is subsequently glucuronidated to SN-38G in the liver and prepared for excretion (42). Gut microbiota then cleave the glucuronic acid group from SN-38G yielding the bioactive SN-38 in the lumen of the gut, exacerbating diarrhea in up to 80% of CRC patients (43). In the case of acetaminophen toxicity, the microbially derived metabolite p-cresol competes with acetaminophen for sulfonation by the host cystolic sulfotransferase enzyme SULT1A1 (31). This causes the shunting of the excess acetaminophen to the liver enzyme CYP450 which converts acetaminophen to the toxic metabolite N-acetyl-p-benzoquinoneimine (NAPQI) (44). Normally, the rate of NAPQI production is not exceeded by NAPQI detoxification through glucuronidation, but as more acetaminophen is shunted into the CYP450 detoxification pathway, free NAPQI will begin to covalently bind to critical cell proteins, causing protein adducts to form in the liver (45). Therefore, elevated serum concentrations of the microbially derived metabolite p-cresol can lead to the accumulation of the hepatotoxic molecule NAPQI, causing liver toxicity and potentially liver failure.

1.2 Contributions

Inspired by the paucity of functional characterization of the human gut microbiome, we made the following contributions to the literature.

In chapter 2, we review how metagenomic, metabolomic, and transcriptomic data were utilized to stratify patient populations based on bilirubin reduction potential, prioritize 20 potential bilirubin reducing strains, and identify a set of 337 candidate bilirubin reductase genes. Untargeted metabolomics data were aggregated and analyzed to produce bilirubin reducer (BR) and non-bilirubin reducers (nonBR) patient labels. These data-driven patient labels were used in conjunction with high throughput metagenomic sequencing data to identify candidate bilirubin-reducing strains which were tested experimentally for the reduction of bilirubin. Finally, we identify a set of 337 candidate bilirubin reductase gene sequences via a custom bioinformatic analysis pipeline.

In chapter 3, we examine sources of hydrogen sulfide (H_2S) production by the human gut microbiome. First, we manually curated known H_2S producing bacterial genes and characterized them into primary, secondary, and erroneous H_2S producer groups based on their mechanism of action. We then conducted a gut microbiome-wide search for these H_2S -producing genes and examine the relative abundance of known and putative H_2S -producing bacterial species across healthy, IBD, adenoma, and CRC patient cohorts. We found that genes responsible for cysteine degradation are ubiquitous in the healthy human gut, and postulate that cysteine degradation may represent the dominant mode of exogenous H_2S production.

In chapter 4, we evaluate the abundance and distribution of azoreductases across the human gut microbiome and compare relative abundances of known and putative azoreducing

species across healthy, IBD, adenoma, and CRC patient cohorts. Starting from primary sequences of known bacterial azoreductases, we constructed profile Hidden Markov Models (HMMs) based on a 35% sequence similarity threshold and searched for these HMMs across 4,644 human gut bacterial genomes. We identified 1,958 putative azoreducing bacterial species and prioritized strains for experimental validation based on health relevance. We confirmed azoreductase activity of three strains of *Fusobacterium nucleatum*, a species that commonly associates with colorectal cancer tumors in the lumen of the gut, indicating capacity for azo bonded prodrug delivery directly to cancerous tissue in the gut.

Taken together, this dissertation describes a methodology to utilize high throughput sequencing and metabolomics data to characterize and predict functions of the human microbiome.

2. Finding bilirubin reductase

2.1 Background

2.1.1 Constitutive bilirubin production and processing

Red blood cells have an approximate lifetime of 120 days after which, they undergo cellular breakdown and processing by macrophages (46). When RBCs are consumed by macrophages, they release large amounts of hemoglobin which require downstream processing. The heme component of hemoglobin is converted to biliverdin via heme oxygenase. Biliverdin is then reduced to unconjugated bilirubin via biliverdin reductase and enters the bloodstream where it binds to albumin. The albumin-bilirubin complex is taken up by the liver for conjugation by UDP-glucoronosyl transferase. Conjugation is a type of first-pass metabolism in which the liver prepares toxic metabolites for excretion in the urine and feces. Conjugated bilirubin is then dumped into the gallbladder via the common bile duct which eventually connects to the gastrointestinal system where it is supposed to be excreted in the feces (47). In the gut, bacteria preform deconjugation of bilirubin via beta-glucoronidases, reversing the work that the liver did. Gut microbiota are also responsible for the reduction of bilirubin to urobilinogen and stercobilinogen, however the gene or genes responsible for bilirubin reduction have not been characterized (48).

2.1.2 Hyperbilirubinemia in neonates and adults

Jaundice is the physical representation of hyperbilirubinemia, and is characterized by yellowing of the skin and whites of the eyes (49). While jaundice is rare in adults, it is exceedingly prevalent in newborns, and one of the most common reasons for infant hospital

readmissions. Jaundice occurs in 60% of full-term neonates and 80% of pre-term neonates, and the very worst cases of neonatal jaundice can lead to acute bilirubin encephalopathy (50). The reason for such high rates of jaundice in newborns is due to: (i) the higher turnover rate of RBCs in neonates and (ii) the underdevelopment of UDP-glucuronosyltransferase 1 (UGT1A1), the gene responsible for bilirubin conjugation in the liver. This causes an influx of bilirubin into the serum which overwhelm the newborn's hepatocytes.

Adults diagnosed with IBD see higher incidences of pigmented gallstones and can have two to four times higher bile bilirubin levels compared to healthy controls (48). In particular, patients with ileal Crohn's disease have elevated bile bilirubin levels due to ileectomy-induced bile acid malabsorption (51). These studies bolster the idea that elevated bile bilirubin levels in Crohn's patients are due to increased enterohepatic cycling of bilirubin.

2.1.3 Potential therapies for neonatal jaundice

The current standard treatment for neonatal hyperbilirubinemia is phototherapy. While phototherapy is an extremely effective treatment, it comes with some downsides. First, photodegradation by-products such as lumirubin have inflammatory side effects. Second, phototherapy requires extended separation from the mother, which has a distressing effect on the newborn. Finally, because neonatal jaundice is the top reason for hospital readmission, it creates an enormous burden on the health care system (50).

Probiotic therapies have been explored and validated for the treatment and prevention of many infant conditions including infantile colic (52,53), *Clostridioides difficile*-associated diarrhea (53,54), antibiotic-associated diarrhea (54,55), and Necrotizing enterocolitis (56–58). A preventative pre- or probiotic treatment for neonatal jaundice could simultaneously

reduce (i) rates of hyperbilirubinemia-induced encephalopathy, (ii) inflammatory side effects of phototherapy, (iii) distress caused by lack of maternal-infant interaction, and (iv) the financial burden placed on the health care system due to jaundice related hospital readmission.

In this work, we leveraged large, publicly available metagenomic and metabolomic datasets to (i) stratify patient populations based on microbial bilirubin reducing potential and (ii) prioritize target strains for experimental validation of bilirubin reduction. Simultaneously, we (iii) identified 337 candidate bilirubin reductase genes using high throughput shotgun metagnomic sequencing data and stratified patient population labels. Finally, we (iv) designed and performed a differential RNA-seq experiment with confirmed bilirubin reducing strains, *Clostridium difficile* cd3, and *Clostridium symbiosum* p11.

2.2 Methods and Results

2.2.1 Stratifying patient populations based on microbial bilirubin-reducing potential

Untargeted metabolomics data from participants of HMP2 (n = 546 samples) and PRISM (n = 218 samples) with corresponding shotgun metagenomic sequencing were used to stratify patients based on their bilirubin reducing potential (Figure 2.1). Counts for bilirubin were compared to the sum of counts for reduced products urobilinogen, stercobilinogen, and their ion complex forms (urobilinogen-Na, urobilinogen-2Na, urobilin-C18-neg, urobilin-FA-C18-neg, stercobilinogen-C18-neg, stercobilinogen-C8-pos, stercobilinogen-HILIC-neg, stercobilinogen-HILIC-pos, stercobilinogen-Na-C18-neg, stercobilin-C18-neg, stercobilin-C8-pos, stercobilin-HILIC-pos, stercobilin-HILIC-neg,

stercobilin-Na-C8-pos, stercobilin-Na-HILIC-neg, stercobilin-[36 adduct]-C18-neg). These measurements are summarized as bil_count and rp_count, respectively. Samples from both HMP2 (59) and PRISM (60) were sorted in descending order according to bil_count and rp_count, resulting in the values bil_rank and rp_rank, respectively. Both bil_rank and rp_rank are non-parametric representations of stool bilirubin and bilirubin reduced product counts, respectively. Next, we calculated the BR_score metric by subtracting the bil_rank and rp_rank for each sample in HMP2 and PRISM. BR_score is normally distributed for both HMP2 and PRISM datasets, and so, z-scores were calculated for each individual sample using $z_i = \frac{BR_{score_i} - mean(BR_{score})}{sd(BR_{score})}$ where z_i is the z-score for sample i, BR_{score_i} is the BR_{score} for sample i, $mean(BR_{score})$ is the mean of all BR_{score} in the specified dataset, and $sd(BR_{score})$ is the standard deviation of BR_{score} for the specified dataset. Finally, we plotted z-scores for all samples in HMP2 and PRISM and designated cutoffs for bilirubin reducing (BR) and non-bilirubin reducing (nonBR) samples at +/- 1.25 z, respectively (Figure 2.1).

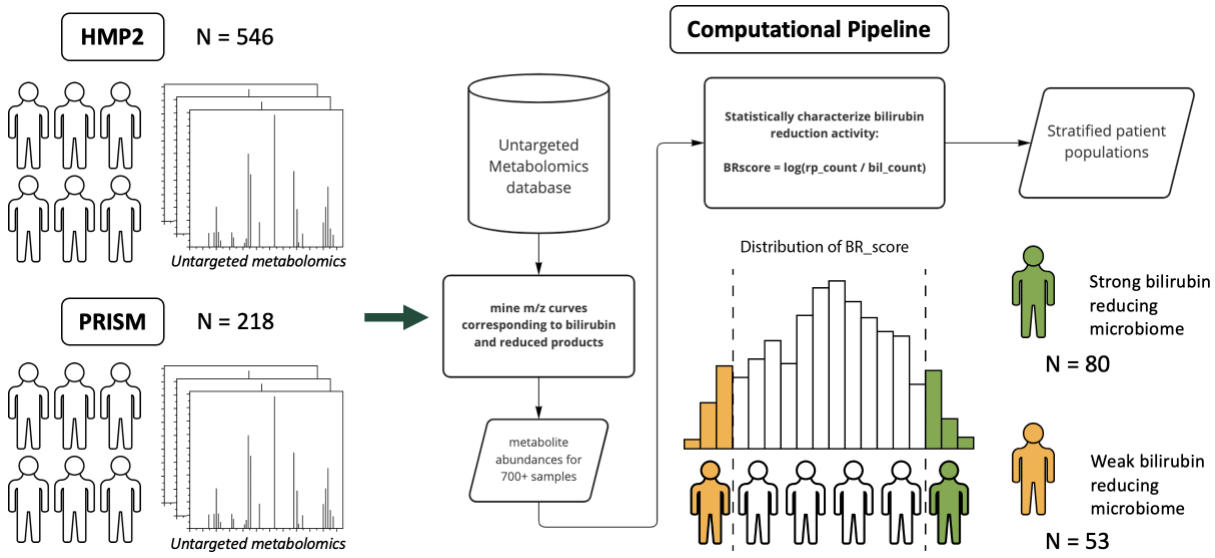

Figure 2.1. Computational and statistical pipeline for patient stratification. Untargeted metabolomics data from HMP2 (n = 546) and PRISM (n = 218) datasets were used to stratify patient populations based on relative abundance of bilirubin and its reduced products, urobilinogen and stercobilinogen. This stratification approach resulted in a group of strong bilirubin reducers (BR) and a group of weak bilirubin reducers (nonBR).

2.2.2 Bacterial strain prioritization

With BR and nonBR labels now created, we utilize corresponding shotgun metagenomic sequencing data to prioritize strains for bilirubin reduction assays. The relative abundance of bacterial species was estimated using MetaPhlAn2 (61) with default parameters for HMP2 and PRISM samples with corresponding metagenomic sequence data. Next, the BR and nonBR labels were used in a differential relative abundance analysis using Maaslin2 (62) with default parameters. Species selected for experimental validation of bilirubin reduction were selected based on three criteria: (i) Adjusted p-value < 0.05 (ii) correlation with BR status > 0.3 and (iii) conditions (i) and (ii) must be satisfied across both HMP2 and PRISM datasets. This resulted in the prioritization of 20 bacterial species for experimental validation of bilirubin reduction via bilirubin reductase (Figure 2.3).

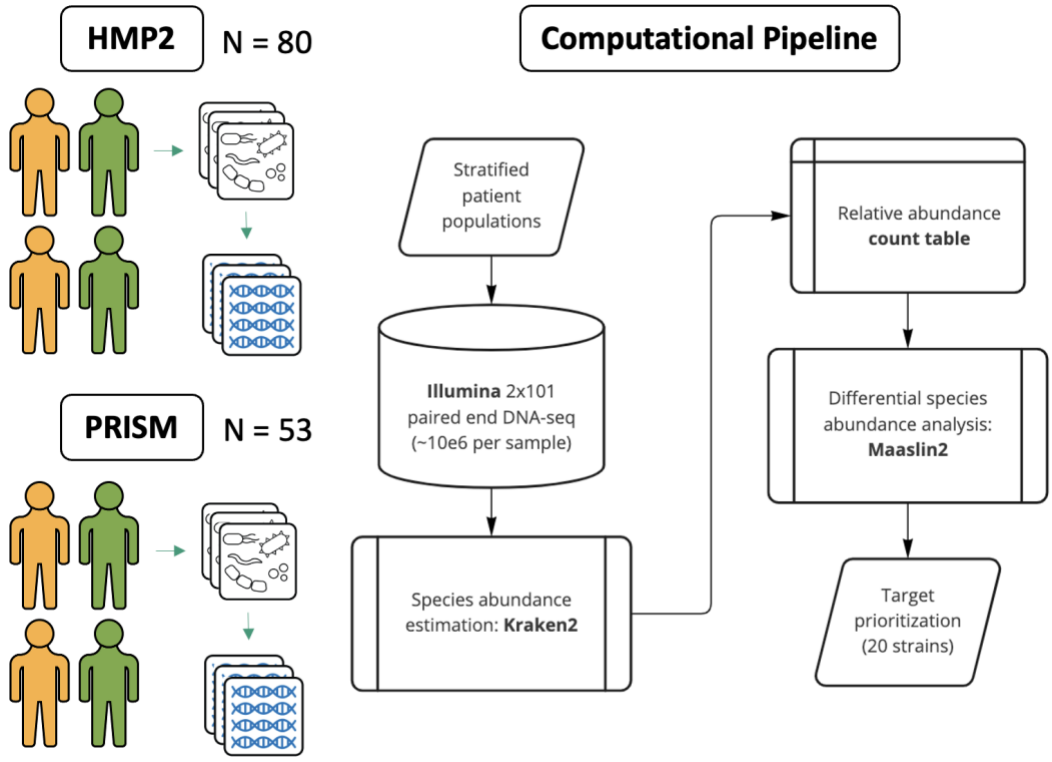

Figure 2.2. Computational pipeline for strain prioritization. Shotgun metagenomic sequencing data for BR and nonBR participants were used to estimate the relative abundance of bacterial species using Kraken 2 (63). The relative the resulting relative abundance count table and BR, nonBR labels were used to prioritize 20 strains for experimental validation of bilirubin reduction. See Figure 2.3 for the results of the differential relative abundance analysis.

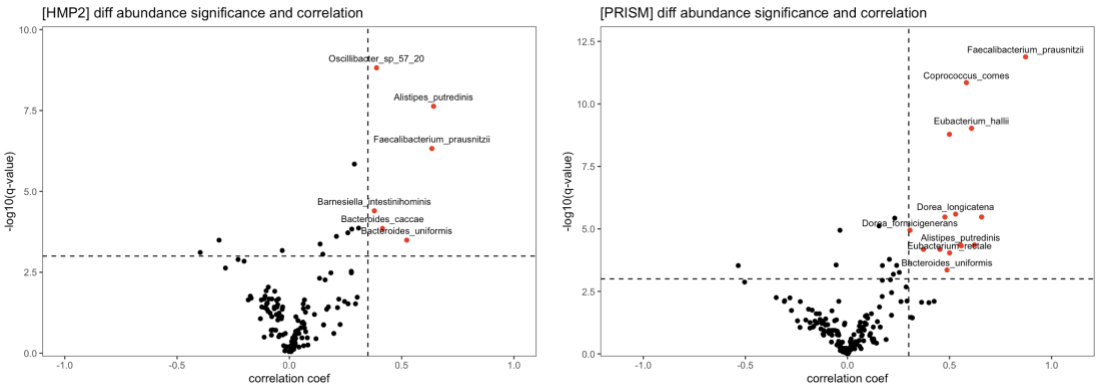

Figure 2.3. Results of strain prioritization for HMP2 and PRISM. Species relative abundances were estimated using MetaPhlAn2 (61). A differential relative abundance analysis was carried out using Maaslin2 (62), and the results are plotted above. Positive

correlation values indicate higher relative abundance in the BR group as compared to the nonBR group.

2.2.3 Bilirubin reductase gene candidate prioritization

The following computational pipeline was designed to identify candidate bilirubin reductase genes present in the human gut microbiome. First, shotgun metagenomic sequencing data from the BR and nonBR labeled samples of HMP2 and PRISM were assembled into contigs using MEGAHIT (64). Next, gene sequences were predicted from those contigs using Prokka (65), and the 2.6×10^6 genes were stored in a custom gene catalog (Figure 2.4). We next used salmon (66) to index the catalog and align back the original metagenomic reads to generate a gene abundance matrix. Finally, we performed differential gene abundance analysis using DESeq2 (67), comparing the abundance of the 2.6×10^6 bacterial genes across BR and nonBR groups. The criteria we used to classify a gene as a candidate bilirubin reductase were: (i) Benjamini-Hochberg adjusted p-value < 0.0001, (ii) positive logFoldChange in BR group, and (iii) both criteria (i) and (ii) must be satisfied for the gene in both the HMP2 and PRISM datasets (Figure 2.5). This resulted in the prioritization of 18,893 candidate bilirubin genes.

2.2.4 Experimental confirmation and comparative genomics analysis

Following bacterial strain prioritization, we performed *in vitro* experimental validation of bilirubin reduction by predicted reducing strains. Prioritized strains were grown in the presence of bilirubin, and a fluorescence assay was performed on the isolates. We measured the fluorescence of urobilinogen, the primary output of bilirubin reduction in 14

strains (Figure 2.6). Putative bilirubin reducing and nonreducing strains were determined from this assay and used as input to a comparative genomics analysis to further narrow down bilirubin reductase candidates. The presence and absence of orthologous genes in the genomes of reducing and non-reducing strains and the presence of a flavin binding motif were used as criteria to narrow down the candidate to one final operon: Bilirubin Reductase Candidate 4 (BRC4) (Figure 2.7). The comparative genomics analysis was completed by collaborators in the Dr. Xiaofang Jiang lab at the National Institutes of Health.

2.2.5 Assessment of BRC4 abundance in BR and nonBR cohorts

We next sought to assess the abundance of BRC4 in BR and nonBR labeled individuals using shotgun metagenomic sequencing data obtained from HMP2 and PRISM studies (Figure 2.8). Illumina paired-end 2x101bp length reads from BR and nonBR labeled samples were processed and quality controlled using fastp (68), and were then aligned to the BRC4 operon using Bowtie2 (69)with default parameters. Read alignment statistics were extracted from sequence alignment map (SAM) files output from Bowtie2 using samtools (70). Alignment statistics were then normalized to counts per million (CPM) for each sample, i, via the following formula:

$$CPM_i = \frac{R_i}{T_i} * C$$

Where R_i is the number of reads aligned to BRC4 for sample *i*, $T_i$is the total number of reads present in sample i, and the scaling factor C = 1x10^6. The abundance of the BRC4 operon was compared across BR and nonBR samples from HMP2 and PRISM datasets. BRC4 is significantly more abundant in BR labeled samples compared to nonBR samples for

both HMP2 ($p = 0.002$) and PRISM ($p = 1.1x10^{-5}$) datasets. Both p-values were calculated using the Wilcoxon Rank Sum Test (71).

2.3 Discussion

Following the strain prioritization analysis outlined in section 2.2.2, we prioritized 20 bacterial species for experimental validation of bilirubin reduction. These strains will be grown anaerobically and tested for their bilirubin reduction ability. With known bilirubin reducing and nonreducing strains identified, we plan to perform a comparative genomics analysis between these sets of bacterial genomes to identify candidate bilirubin reductase genes. In a parallel computational workflow outlined in section 2.2.3, we identified 18,893 candidate bilirubin reductase genes. With the results of these two parallel prioritization pipelines in hand, we planned to further narrow down the set of candidate bilirubin reductase genes based on: (i) Enzyme Commission (EC) number relevancy, (ii) synteny within the bacterial genome, and (iii) containment in confirmed bilirubin reducing strains and exclusion from confirmed non-bilirubin reducing strains.

Finally, a comparative genomics analysis was conducted by collaborators working in the Xiaofang Jiang lab at the National Institutes of Health, which narrowed the bilirubin reductase candidate list down to one three-gene operon: BRC4 (Figure 2.7). This operon encodes three genes: a transcriptional regulator (BRC4X), an NADH oxidase (BRC4Y), and a flavodoxin protein (BRC4Z) (Figure 2.7). Finally, we show that the BRC4 operon is significantly higher in abundance in BR samples compared to nonBR samples across both HMP2 and PRISM datasets (Figure 2.8). Taken together, these findings further support the hypothesis that BRC4 encodes a bilirubin reductase.

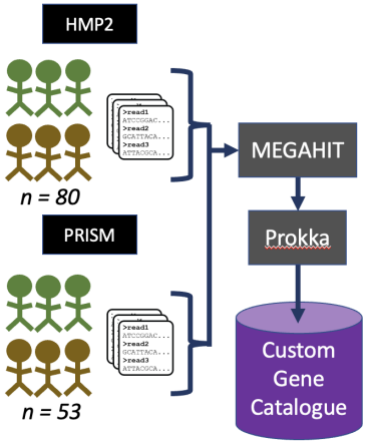

Figure 2.4. Custom gene catalogue generation pipeline. Raw metagenomic sequencing data for BR and nonBR samples in HMP2 and PRISM were assembled into contigs using MEGAHIT (64). Gene sequences from those contigs were predicted using Prokka (65), which then served as the gene catalog for the differential gene abundance analysis.

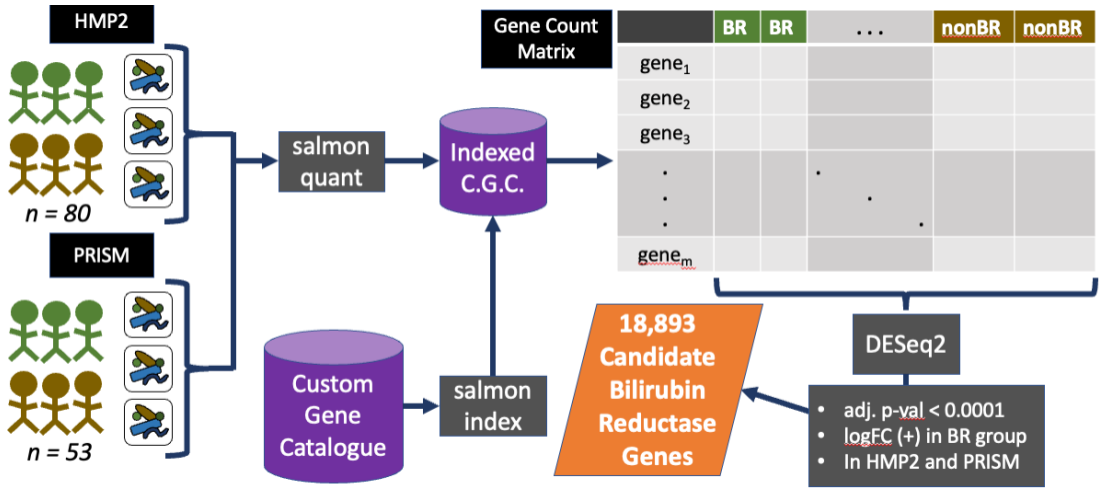

Figure 2.5 Differential gene abundance analysis. Raw metagenomic sequence data were used to quantify the abundances of genes present in the custom gene catalog using salmon (66). The resulting gene abundances and BR nonBR patient labels were used in a differential gene abundance analysis using DESeq2 (67). Candidate bilirubin reductase genes were identified based on the criteria present in the bottom right of the above figure.

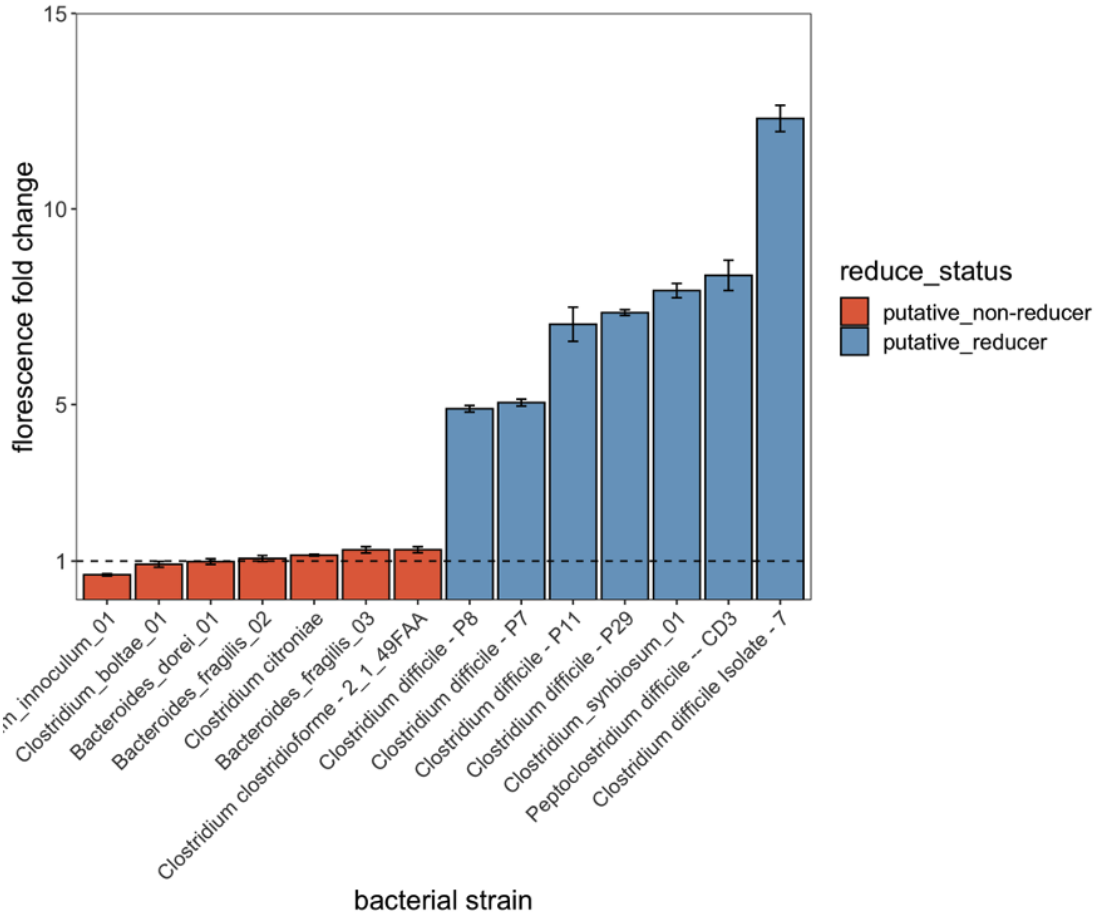

Figure 2.6. Bilirubin reduction fluorescence assay. Shown above are the results of the bilirubin reduction assay for 14 different strains of the *Clostridium*, *Peptoclostridium*, and *Bacteroides* genera. The y-axis shows the fold change in fluorescence compared to the blank control for these strains. Putative bilirubin reducing strains are labeled in blue and putative non-reducing strains are labeled in red. These genome sequences of these strains were used as input to a comparative genomics analysis to further narrow down candidate bilirubin reductase genes.

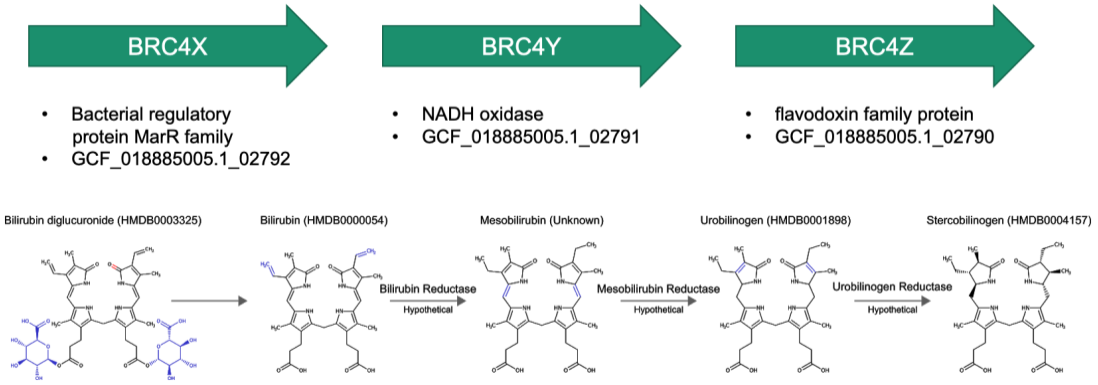

Figure 2.7. Bilirubin Reductase Candidate 4. (Top) orientation of the Bilirubin Reductase Candidate 4 (BRC4) operon in the *Clostridium difficile* cd3 genome. (Bottom) Proposed mechanism for deconjugation and subsequent carbonyl bond reduction of bilirubin to reduced products Urobilinogen and Stercobilinogen. Mesobilirubin is a hypothesized intermediate molecule that is not recovered in any *in vitro* bilirubin reduction assay thus described.

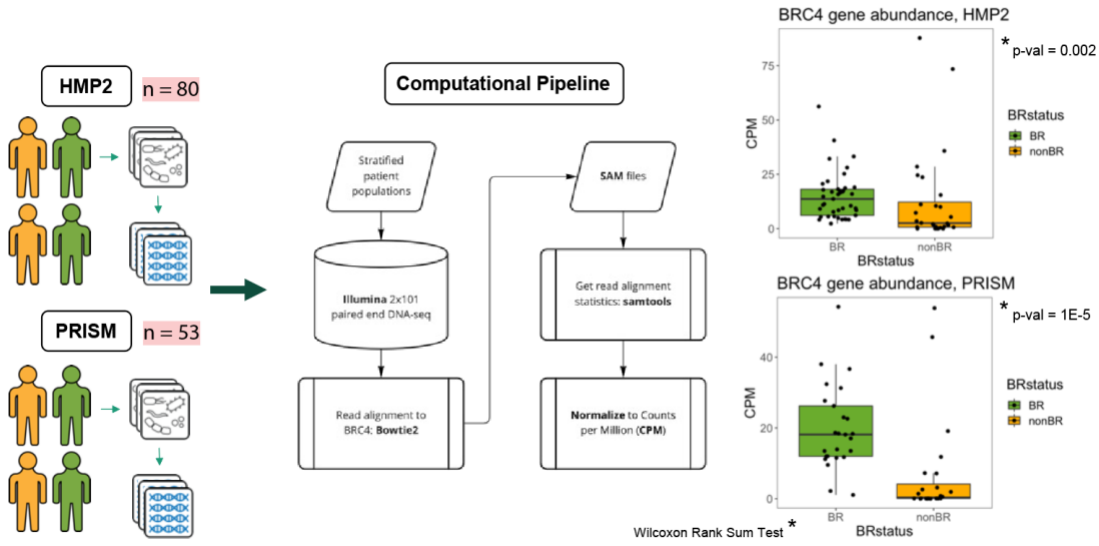

Figure 2.8. Title. (Left) Shotgun metagenomic sequencing data taken from BR and nonBR samples were processed and prepared for alignment to the BRC4 operon. (Middle) We performed alignment of cleaned and processed Illumina paired end 2x101bp reads to the BRC4 operon using Bowtie2. Read alignment statistics were extracted from the Sequence Alignment Files (SAM files) output using samtools. Finally, the read alignment statistics were normalized to counts per million (CPM) by dividing reads aligned to BRC4 by the total number of reads per sample. These counts were then multiplied by a scaling factor = 1x10^6 to obtain CPM. (Right). The abundance of the BRC4 operon in HMP2 and PRISM samples. BRC4 is significantly higher in abundance in BR samples for both HMP2 (p = 0.002) and PRISM (p = 1x10^-5) datasets.

3. The capacity to produce hydrogen sulfide (H_2S) via cysteine degradation is ubiquitous in the human gut microbiome

3.1 Background

Hydrogen sulfide (H_2S) is a consequential molecule produced by the gut microbiota with pleiotropic effects on human physiology. It is one of the three physiological gasotransmitters, along with carbon monoxide and nitric oxide, and is produced endogenously in many tissues including, but not limited to, the brain, heart, and liver (72). Endogenous H_2S production occurs via the enzymes cystathionine beta-synthase (cbs), cystathionine gamma-lyase (cse) and 3-mercaptopyruvate sulfur transferase (3mst) (73). cbs, cse and 3-mst are tightly regulated pyridoxal-5′-phosphate (PLP)-dependent enzymes and produce H_2S primarily from the degradation of cysteine (74) (Figure 3.1B). H_2S produced by these enzymes plays a litany of physiological roles including suppression of oxidative stress in the brain, regulation of blood pressure through vasodilation, and protection of hepatic stellate cells from cirrhosis in the liver (75). As a result, abnormally low endogenous levels of H_2S are hypothesized to be an underlying cause of peripheral artery disease, and efforts have been made to measure serum levels of H2S quickly and non-invasively as a proxy for early detection of peripheral artery disease (76).

Microbes in the gastrointestinal tract also contribute to H_2S production in humans. A majority of the microbially produced H_2S originates in the colon, where estimates of luminal concentrations of H_2S range from 0.3 to 3.4 mM (27). The serum concentration of H_2S in healthy individuals is difficult to measure but is estimated to range from 34.0 to 36.4 μM 8/15/2022 2:44:00 PM. H_2S readily diffuses across the intestinal epithelium and can enter

circulation influencing host physiology (77). Excessive production of H_2S by gut microbes has been linked with decreased mucosal integrity through reduction of mucosal disulfide bonds (78), inhibition of colonocyte butyrate oxidation via cytochrome-c inhibition (Gibson et al., 1988), and genotoxicity (77) (Figure 3.1C). A prime example of the gut microbiome affecting gut health is association of Fusobacterium nucleatum, a known H_2S producer with colonic tumors (79) (Figure 3.1C).

While the mammalian pathways of H_2S production have been well studied, the contribution of gut-microbial H_2S production to circulating H_2S levels and the subsequent systemic effects on human physiology are largely unknown. The first step toward a better understanding of the effects of H_2S on human physiology is to identify which microbial species are responsible for H_2S production. There are two major sources of H_2S production in the human gut microbiota, dissimilatory sulfate reduction (DSR) and the degradation of the sulfur-containing amino acids cysteine and methionine (80). We must note that sulfate is first reduced to sulfite before H_2S is produced, however, we refer to this process as sulfate reduction for the remainder of this work.

In the literature, H_2S production is often attributed to the well-characterized dissimilatory sulfate reduction pathway (75). Common representatives of sulfate-reducing bacteria (SRB) are found in the phylum Desulfobacterota (recently reclassified from the class Deltaproteobacteria) with *Desulfovibrio* spp. and *Bilophila wadsworthia* being the most abundant representatives in the human gut (81,82). Sulfate and sulfite are used by SRB as terminal electron acceptors for anaerobic respiration (83–85). While SRB are prevalent in human populations, their relative abundances are generally very low and are dependent on ecological interactions with other hydrogenotrophs, such as methanogens and acetogens

(81,84,85). Methane (CH_4) is primarily produced by the methanogen Methanobrevibacter smithii (86) and is one of the primary gases present in mammalian flatus. Sulfate-reducing bacteria and methanogens have been historically considered mutually exclusive in microbial communities due to the competition for hydrogen (83). However, experiments carried out on human flatus have shown that both H_2S and CH_4production occurs simultaneously in some individuals, seemingly contradicting the notion that methanogens and sulfate-reducing bacteria cannot co-exist (27). Unlike the comprehensively characterized pathways for dissimilatory sulfate reduction, the species of the gut microbiome responsible for H_2S production via degradation of sulfurcontaining amino acids (cysteine and methionine) have not been comprehensively characterized. Gut microbial involvement in amino acid fermentation has garnered recent attention, as many physiologically relevant downstream metabolites are produced by gut microbial degradation of amino acids (87) (Figure 3.1A). Depending on dietary intake, a pool of sulfur-containing amino acids is available for fermentation by gut microbiota (88). Various studies have demonstrated that cysteine supplementation leads to far more H_2S production than inorganic sulfate supplementation underscoring the comparative importance of the cysteine degradation pathway in total H_2S production (84,85,89). It is important to delineate between H_2S produced via dissimilatory sulfate reduction and H2S produced via cysteine degradation because different approaches are necessary to modulate these two sources of H_2S production. Because of the poor annotation of the genes which produce H_2S via cysteine degradation across species of the gut microbiome, the relative contributions of cysteine-degradation and sulfate reduction to overall exogenous H_2S production are unclear. To address this gap, we designed a bioinformatic approach to first identify putative cysteine-degrading bacteria in the human gut

microbiome and then compared the relative abundances of putative cysteine-degrading bacteria and sulfate-reducing bacteria across metagenomic data from Inflammatory Bowel Disease (IBD), colorectal cancer (CRC), and healthy cohorts (Figure 3.5).

3.1.1 Endogenous H_2S production

3.2 Results

To identify species capable of H_2S production via cysteine degradation in the human gut microbiome, we curated profile Hidden Markov Models (pHMMs) of enzymes with experimental evidence of H_2S production via cysteine-degradation and classified them into primary, secondary, and erroneous producers of H_2S. Enzymes which produce H_2S via cysteine degradation as their primary function are labeled "primary" (dcyD, yhaM, mgl, sseA) and enzymes that also participate in separate mechanism(s), such as the transsulfuration pathway and maltose regulon modulation, have been labeled "secondary" (metC, malY, cysK, cysM, mccB). Enzymes which have a well-defined function other than H_2S production via cysteine-degradation are labeled "erroneous" (tnaA, iscS, mccA). Please see Appendix 1 for more information.

3.2.1 Cysteine-degrading genes are widely distributed in the human gut microbiome

We performed a homolog search for these H_2S-producing enzymes across 4,644 species in the Unified Human Gastrointestinal Genome (UHGG) collection (90) using HMMER (91) (Figure 3.2 and Figure 3.5A). This collection comprises 4,644 non-redundant genome sequences from species representatives generated by clustering 204,938 genome sequences from bacteria known to inhabit the human gut. Of the representative UHGG

species, 44.1% (2,046/4,644) contain one or more primary cysteinede grading gene, 42.0% (1,951/4,644) contain one or more secondary cysteine-degrading gene, and 1.1% (5/4,644) contain one or more erroneous cysteine-degrading gene. Aside from known cysteine-degrading bacterial species compiled in the manual curation step, an additional 3,065 species from the UHGG catalog were found to contain one or more cysteine degrading genes (Figure 3.2).

The prevalence and relative abundance of putative cysteine degrading bacteria and sulfate-reducing bacteria was calculated for 10,700 metagenomic samples from healthy, IBD, CRC, and adenoma cohorts (59,60,92,93). Among the 6,632 healthy subjects, there is a markedly higher relative abundance of putative primary and secondary cysteine-degrading bacteria compared to sulfate-reducing bacteria ($p < 2.2e\text{-}16$, two-sided Wilcoxon Rank Sum Test) (Figure 3.3). This suggests that cysteine degradation may contribute considerably to H_2S production for the average healthy person. Cysteine-degrading genes are also widespread in healthy populations with 100% of the 6,623 healthy subjects containing at least one putative cysteine-degrading bacteria in their gut microbiome (Figure 3.3).

3.2.2 Increased relative abundance of putative hydrogen sulfide-producing bacteria in the Inflammatory Bowel Disease and colorectal cancer gut microbiomes

We assessed the relative abundance of putative cysteine degrading bacteria and sulfate-reducing bacteria in individuals with the two most common clinical manifestations of Inflammatory Bowel Disease (IBD), Crohn’s disease and ulcerative colitis, colorectal cancer (CRC) and healthy controls (94). Putative primary and secondary cysteine-degrading bacteria are significantly more abundant than sulfate-reducing bacteria across IBD and CRC

populations from metagenomic samples obtained from curatedMetagenomicData (93), the Integrative Human Microbiome Project 2 (HMP2) (59), PRISM (60,92) (all $p < 2.2 \times 10{-}16$) (Figures 3A–D).

Putative primary cysteine-degrading bacteria are significantly more abundant in CRC than in the control groups ($W = 123{,}784$, $p = 7.4 \times 10{-}11$) (Figure 3A) while putative secondary cysteinedegrading bacteria were found to be significantly less abundant in CRC compared to healthy controls ($W = 79{,}734$; $p = 1.2 \times 10{-}6$). A similar trend follows for ulcerative colitis and Crohn's disease compared to healthy controls. Putative primary cysteine degrading bacteria are significantly more abundant in IBD when compared to healthy controls and putative secondary cysteine degrading bacteria are significantly less abundant than in healthy controls (Figures 3B,C). Within pediatric Crohn's disease, there is no apparent difference in the abundance of putative primary and secondary cysteine-degrading bacteria (Figure 3D).

3.2.3 Primary and secondary cysteine-degrading genes are actively expressed in the human gut microbiome

To confirm in vivo transcription of cysteine-degrading genes and sulfate-reducing genes in the human gut microbiome, we analyzed 736 metatranscriptomic sequencing runs from 318 individuals across two studies: (1) The Health Professionals Follow-up Study (number of individuals = 308, number of samples = 677) (95) and (2) (96) (number of individuals = 10, number of samples = 59). Both studies took multiple samples from participants over the course of time and/or dietary intervention, hence the distinction between number of individuals and number of samples. Our analysis revealed that 86.5% (637/736) of samples show expression of at least one primary cysteine-degrading gene, 89.7% (660/736)

of samples showed expression of at least one secondary cysteine-degrading gene, and 84.1% (619/736) of samples showed expression of dissimilatory sulfate reduction genes dsrA and dsrB (Figure 4 and Supplementary Figure 4). yhaM and mgl are the most actively transcribed primary cysteine-degrading genes with malY and cysK being the most actively transcribed secondary cysteine-degrading genes. The erroneous cysteinedegrading genes tnaA, iscS, and mccA are considerably less transcribed across healthy human gut microbiomes (Figure 4). These results suggest that primary and secondary cysteine degradation could be prominent pathways of H_2S production in some individuals.

3.2.4 Core dissimilatory sulfate reduction genes and methanogenesis genes are co-expressed in vivo

Previously, in vitro assays have indicated that methanogens and sulfate-reducing bacteria compete for hydrogen and may thus mutually exclude one another (Gibson et al., 1988). However, through analysis of 736 metatranscriptomic samples obtained from 318 individuals across two studies (95,96), we observed that core genes involved in dissimilatory sulfate reduction and methanogenesis are simultaneously expressed in 25.8% (175/677) of samples from the HPFS study and in 11.9% (7/59) of samples from the David et al. (2014) study (Supplementary Figure 4). This suggests that the mutual exclusivity of sulfate-reducing bacteria and methane-producing bacteria observed in vitro does not necessarily apply to the complex biogeography of the gut.

3.3 Discussion

Due to its role as a mammalian gasotransmitter, H_2S plays important roles in maintaining physiological homeostasis. However, H_2S may also cause deleterious effects in a

concentration-dependent manner. Therefore, it is of great importance to understand the sources of exogenous H_2S production in the gut in order to tease out the links between H_2S and human physiology. The source of gut microbial H_2S production is often attributed to dissimilatory sulfate reduction, with far less attention given to H_2S production via the degradation of the sulfur-containing amino acid cysteine. In fact, there has not been a microbiome-wide annotation of the potential for H_2S production via cysteine degradation. The systematic annotation we performed in this study expands our understanding of which species can potentially produce H_2S in the gut, the majority of which have not been previously reported to have the capability for H_2S production. Our analysis of shotgun sequenced metagenomic data from 10,700 metagenomic samples revealed that putative cysteine-degrading bacteria are ubiquitous inhabitants of the human gut microbiome and are present at significantly higher relative abundance than sulfate-reducing bacteria. Furthermore, our analysis of 736 metatranscriptomic samples from 318 healthy individuals demonstrates that primary and secondary cysteine-degrading genes are, in fact, actively expressed in the gut. These results suggest that primary and secondary cysteine degradation could be prominent pathways of H_2S production in some individuals. Therefore, cysteine degradation is an important aspect to consider when designing studies to assess the effects of H_2S on human health or modulate gut microbial H_2S production.

We also explored the relative abundance of putative cysteine degrading bacteria in IBD and CRC to understand whether these bacteria could contribute to or promote disease progression. We found that primary putative cysteine-degrading bacteria are significantly more abundant in CRC samples than in healthy controls. While relative abundances of sulfate-reducing bacteria are modestly higher in CRC compared to healthy controls, primary

putative cysteine-degrading bacteria are far more abundant. This finding corroborates previous studies linking H_2S and the progression of CRC (79) and highlights the need to identify the dominant source of H_2S in the CRC gut. Importantly, it remains to be elucidated whether or not this difference in relative abundance translates to higher production of H_2S via cysteine degradation.

Prior studies suggested that methanogens and sulfate-reducing bacteria are mutually exclusive, potentially due to their competition for hydrogen. However, subsequent studies have reported the presence of both CH_4 and H_2S in the human flatus (27), seemingly contradicting this notion of mutual exclusivity of CH_4 and H_2S-producing bacteria. To resolve this discrepancy, we examined the transcriptional cooccurrence of methanogens and sulfate-reducing bacteria, and cysteine-degrading bacteria in the human gut and found the co-occurrence of all three pathways. This discrepancy between in vitro experiments and in vivo observations could be explained by the complex biogeography of the gut in which methanogens and sulfate-reducing bacteria occupy distinct niches or from H_2S production via cysteine degradation.

The primary limitation of this study is the lack of experimental confirmation of primary, secondary, and erroneous putative cysteine-degrading bacteria. This could be addressed by performing a screen for H_2S production via cysteine degradation for all culturable strains that we have identified as putative cysteine-degrading bacteria. There are many reactions in which H_2S is formed as an intermediate, such as assimilatory sulfate reduction, however, these reactions do not result in significant production of H_2S and are thus not relevant to total H_2S production by the gut microbiome. Therefore, we limited our search for H_2S -producing bacteria to pathways in which H_2S was the endpoint, or byproduct, and

not just an intermediate of the pathway. Our search identified the genes for dissimilatory sulfate reduction in Eggerthella and Gordonibacter species. We have included these species as sulfate-reducing bacteria though there is little evidence to suggest that these species are true sulfate reducers (97,98). Experimental validation of these claims is necessary to confirm Eggerethella spp. and Gordonibacter spp. as non-sulfate-reducing bacteria. We also note that our search for H_2S-producing genes included only the 4,644 representative genomes in UHGG. The full UHGG collection contains 204,938 non-redundant genomes with core and accessory gene information that may contain other putative H_2S-producing sub-species that we did not analyze. Another potential shortcoming of this analysis is the overrepresentation of western countries in the data pool used. An expanded set of samples would be required to claim that primary and secondary putative cysteine-degrading bacteria are globally prevalent in the human gut microbiome. Finally, we note that sulfate-reducing bacteria may be mucosally associated and present at low relative abundances which could mean that stool metagenomics may underestimate the true abundance of sulfate-reducing bacteria in the human gut.

3.4 Methods

3.4.1 Curation of Profile Hidden Markov Models Involved in Cysteine-Degradation and Sulfate-Reduction

We performed a literature search to identify genes involved in H2S production via dissimilatory sulfate reduction and sulfur-containing amino acid degradation by gut bacteria (29,80,99–102). The pHMMs corresponding to the gene families responsible for H2S production were obtained from TIGRFAM or HAMAP. If neither a TIGRFAM nor HAMAP

pHMM was available for a given gene or the profile was too broadly defined, we opted to construct a custom pHMM to represent the gene of interest. Custom pHMMs were constructed by 1. concatenating amino acid sequences pertaining to the gene of interest 2. performing a multiple sequence alignment over these amino acid sequences using MUSCLE (103) 3. constructing the pHMMs from the multiple sequence alignments using hmmbuild from the HMMER tool suite (91). This was done to avoid identifying spurious homolog hits in the subsequent search step.

3.4.2 Search for Putative Hydrogen Sulfide Producing Bacteria in the Human Gut

The pHMMs of H2S producing genes were searched against 4,644 genome sequences from UHGG (90) using the hmmscan method from hmmer v3.1 (91) (Supplementary Figure 1A). Hits were filtered based on a conservative E-value threshold (E-value < $1 \times 10{-}110$) and an additional bit score threshold was applied for TIGRFAM pHMMs to avoid calling spurious hits. Hits to HAMAP and custom pHMMs were only filtered based on the E-value threshold. Next, the bacterial genomes receiving hits were categorized into putative primary, secondary, and erroneous cysteine degraders based on the known mechanisms of the genes used in the search space. Genomes receiving hits to the dsrAB operon were labeled as sulfate-reducing bacteria. Please see Supplementary Note 1 for a detailed description of the primary, secondary, and erroneous classification scheme. Primary, secondary, and erroneous putative cysteine-degrading bacteria across UHGG were then visualized by uploading a taxonomic tree in newick tree format to the iTOL (104) web interface (Figure 2). Gene containments for each of the UHGG genomes were converted to the EXTERNALSHAPE file

format specified by iTOL documentation and uploaded to the iTOL visualization file from the previous step.

3.4.3 Calculating relative abundances with Kraken 2

Raw sequencing reads for samples from curatedMetagenomicData, HMP2, PRISM, and Lewis et al. (2015) were downloaded and extracted with NCBI's SRA toolkit v2.10.9 (105). Quality control and adapter trimming of the fastq sequence files were done with the Trim Galore wrapper v0.6.6 (106). To remove potential human contaminants, quality trimmed reads were screened against the human genome (hg19) with Bowtie2 v2.4.2 (69). Taxonomy profiling of the cleaned metagenomic reads were generated using Kraken 2 (2.0.8-beta) (63) to estimate the relative abundance of bacterial species present in each dataset. These relative abundances are then processed and plotted in Figure 3.

3.4.4 Transcriptomic Analysis of Hydrogen Sulfide Producing Genes and Methane Producing Genes

We sought to confirm the active expression of H2S producing genes and CH4 producing genes alongside the existing genomic evidence presented using metatranscriptomic data from David et al. 2014 (96) and the Health Professionals Follow-up Study (95). Confirming the expression of H2S producing and CH4 producing genes involved the following steps: 1. Metadata for samples were downloaded from the SRA run selector.2 2. Raw sequencing data were downloaded using fasterq-dump from the SRA toolkit version 2.10.9 (105). 3. Manually curated H2S producing genes and CH4 producing genes were given as input to salmon index (66). 4. Raw RNAseq data were then quantified against the manually curated gene sequence database using the salmon quant command with the

validateMappings option on for better performance. The raw counts of reads mapped per gene were normalized to TPM values for downstream analysis. The threshold for considering an H2S gene “expressed” was TPM > = 10. A sample was said to be “methane producing” if ≥80% of the 16 genes involved in the methanogenesis pathway recruited one or more read mapping. These genes are listed in the x-axis of Supplementary Figure 3. The results were then parsed with custom shell and R scripts and visualized in Figure 4 and Supplementary Figure 3 using the R package ggplot2 (107). All scripts and workflow is available at https://github.com/dombraccia/H2S.

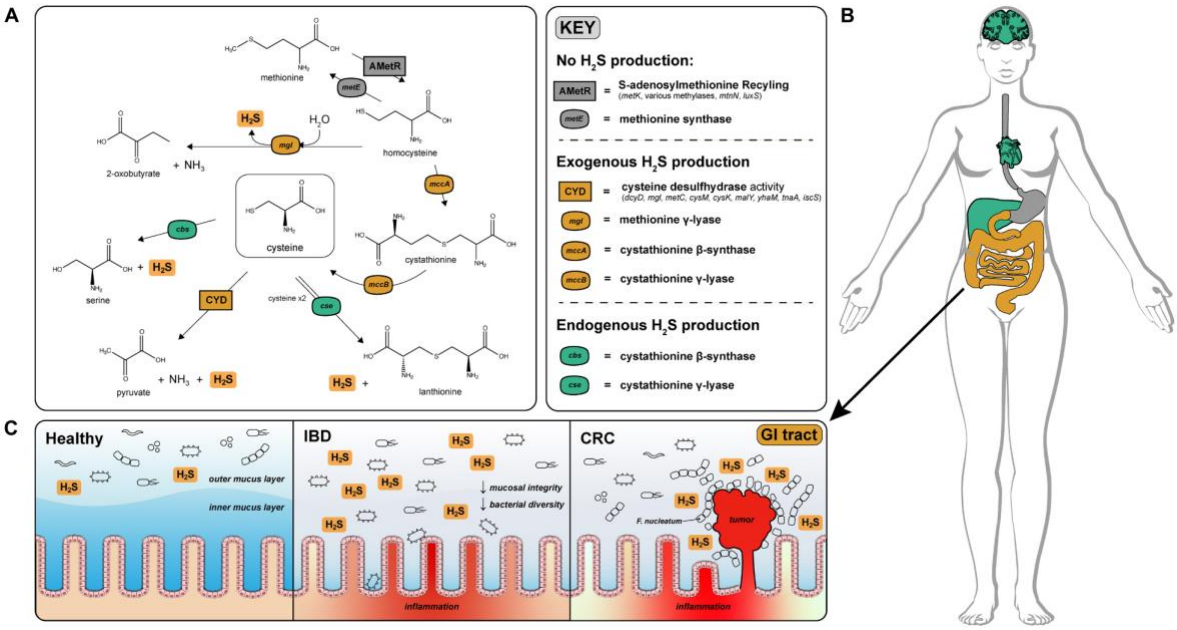

Figure 3.1. H_2S production via cysteine degradation in the human gut microbiome. (A) Pathways of H2S production via cysteine degradation in the human gut microbiome. Pathways with labels ending in "activity" refer to a set of genes that convert cysteine to the indicated products. Cysteine desulfurase activity (CYD) = (dcyD, mgl, metC, cysM, cysK, malY, yhaM, tnaA, iscS). AMetR = AdoMet recycling present in Bacillus subtilis (metK, various methylases, mtnN, luxS) (108) (B) Visualization of H2S production across human tissues (image obtained from Papatheodorou et al., 2020 (109)). H2S is produced endogenously in the brain, liver, and heart via cystathionine gamma-lyase and cystathionine beta-synthase and is tightly regulated to avoid toxic effects of H2S overproduction. Color coded organs refer to the type of H2S production active in those organs. (C) Physiological effects of H2S on the healthy, IBD and CRC gut. H2S contributes to the degradation of the protective mucosal barrier which could cause or exacerbate inflammation and infection by opportunistic species for patients diagnosed with IBD. In CRC, Fusobacterium nucleatum is closely associated with colonic tumors and are well known H2S producers (79).

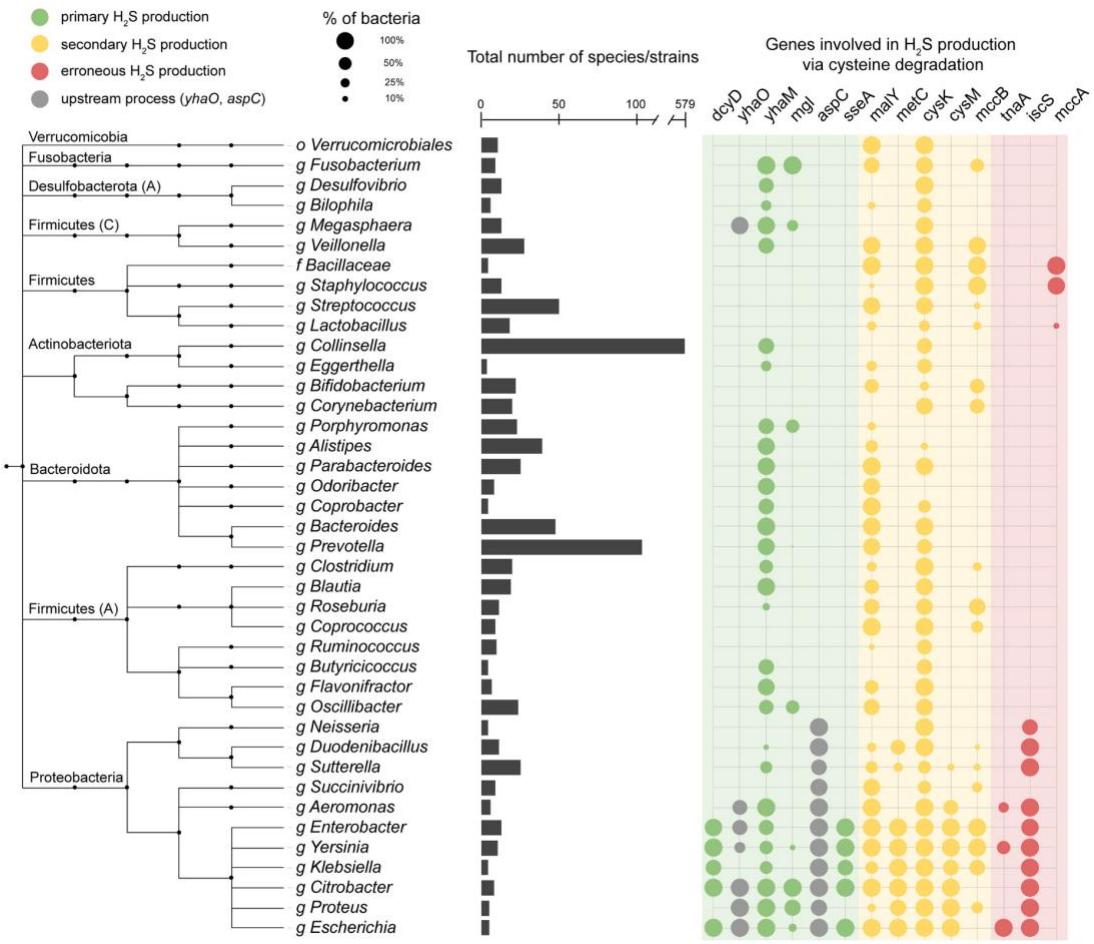

Figure 3.2. Presence of cysteine-degrading genes across the human gut microbiome. Important genera of the human gut microbiome and the presence/absence of cysteine-degrading genes in each clade. The taxonomic tree is obtained from the Unified Human Gastrointestinal Genome collection (UHGG) (90) which is built on the Genome Taxonomy Database (GTDB) (110). Phyla names are annotated on the left side. Phyla followed by a capital letter, e.g., Desulfobacterota (A), indicate a novel phyla classified by the GTDB-tk (110). The bar chart in the center of the figure represents the number of species contained under each genus or higher clade. The color of the circles on the right indicates whether the gene is a primary (green), secondary (yellow) or erroneous (red) producer of H2S (Appendix 1). The circles on the right side represent the number of species in each clade that contain hits to the genes specified. Nodes collapsed at levels higher than genus are because all genomes in that clade contain the same combination of genes reported in the grid on the right.

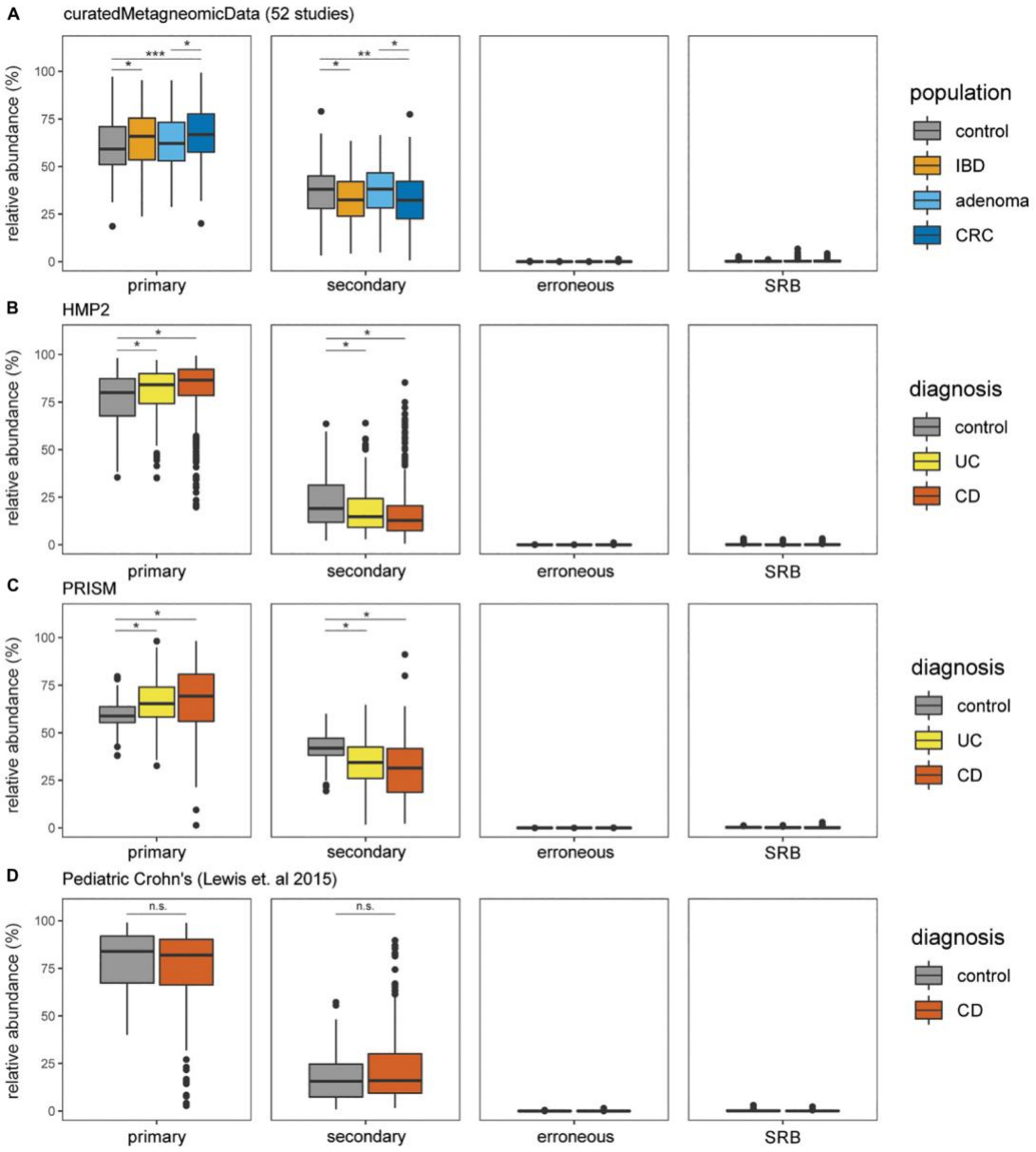

Figure 3.3. Putative primary and secondary cysteine-degrading bacteria are more prevalent than (SRB) among individuals with IBD, CRC, and healthy controls. Relative abundances of putative cysteine-degrading bacteria and sulfate-reducing bacteria across healthy and diseased populations. Relative abundances were calculated using Kraken2 (63). (A) Data obtained from curatedMetagenomicData (93). Number of samples per disease category: control = 560, CRC = 352, adenoma = 143, IBD = 148. (B) Data obtained from HMP2 (59). Number of samples per disease category: non-IBD = 359, ulcerative colitis (UC) = 367, Crohn's disease (CD) = 591. (C) Data obtained from PRISM (60). Number of samples per disease category: control = 56, UC = 76, CD = 88. (D) Data obtained from a study on Pediatric Crohn's Disease (92). Number of samples per disease category: control = 26, CD = 86.

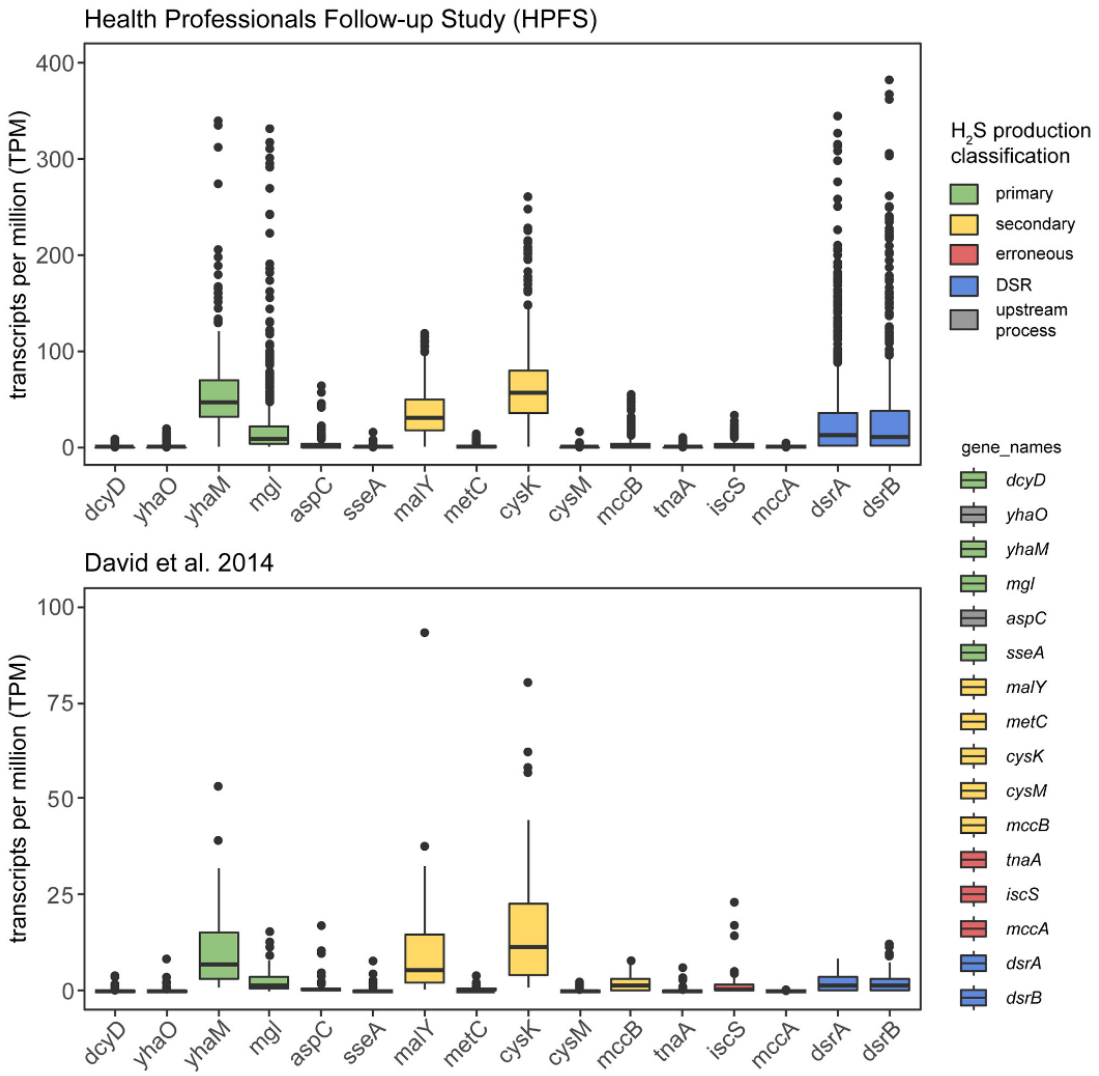

Figure 3.4. Cysteine-degrading genes are actively expressed in healthy adults. This analysis confirms that the H2S producing genes considered in this work are actively expressed in healthy adults. The y-axis displays TPM counts for each gene involved in H2S production via both cysteine degradation and dissimilatory sulfate reduction. The legend on the right indicates the mode of H2S production for each of the genes examined. Metatranscriptomic reads from HPFS (number of individuals = 308, number of samples = 677) (95,96) (number of individuals = 10, number of samples = 59) (96) were aligned to gene hit sequences identified in the homolog search step using salmon (66) and visualized by the ggplot2 (107) package in R (Figure 3.5). Certain primary and secondary H2S producing genes are actively produced by gut bacteria in healthy adults and erroneous producers of H2S appear to be less actively transcribed.

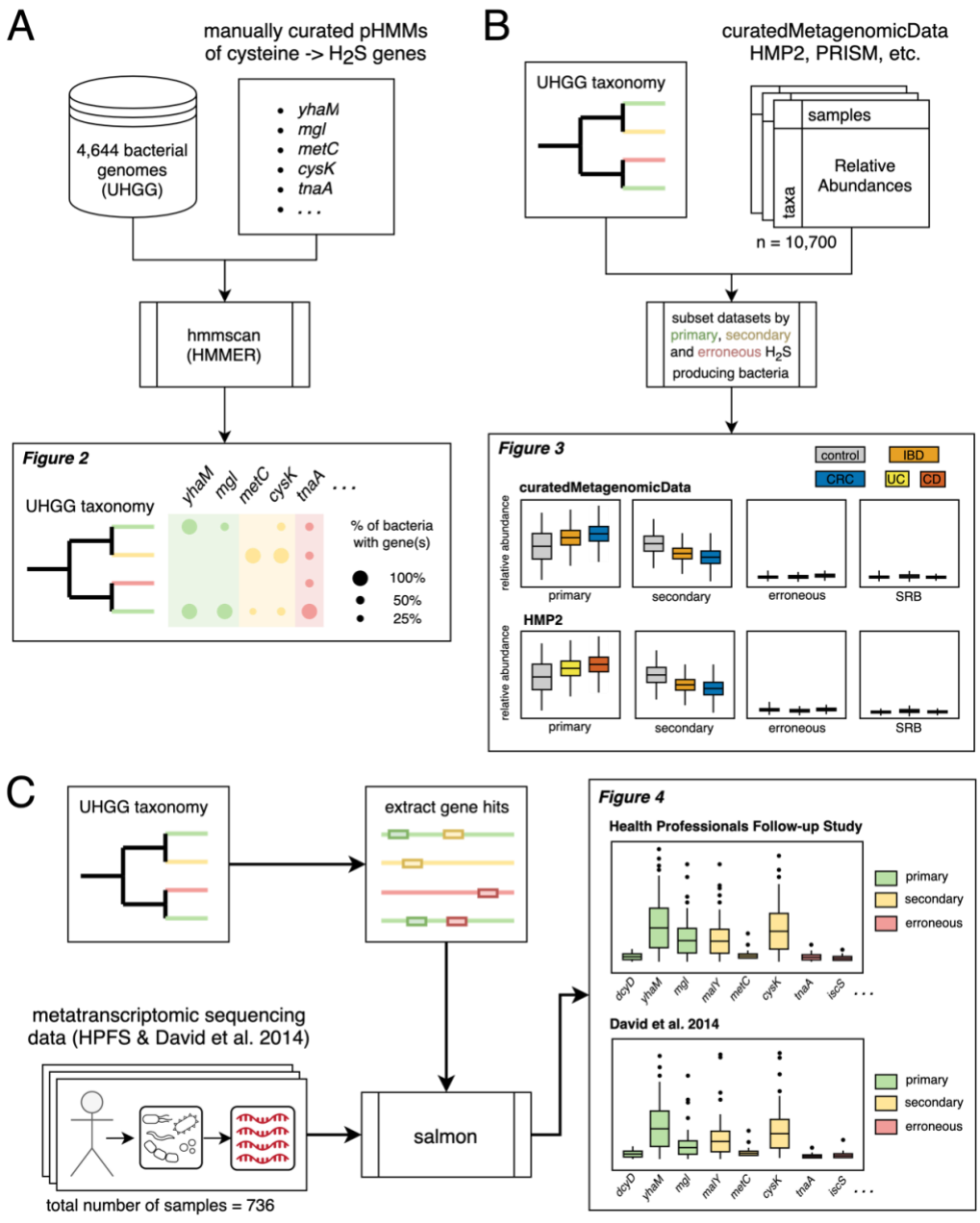

Figure 3.5. Computational workflow. (A) Representative bacterial genomes from the United Human Gastrointestinal Genomes collection (UHGG) (90) were downloaded and profile hidden markov models (pHMMs) representing cysteine-degrading genes were manually curated after extensive literature search (Appendix 1). pHMMs were searched across the UHGG database using hmmscan (HMMER) (91) and gene containments are reported in Figure 3.2. **(B)** Relative abundances of putative primary, secondary and erroneous cysteine-degrading bacteria were compared across healthy, IBD and CRC individuals and reported in Figure 3.3. **(C)** Metatranscriptomic sequencing reads from HPFS (number of individuals = 308, number of samples = 677) (95,96) number of individuals = 10, number of samples = 59) were aligned and quantified against UHGG gene hits using salmon (66) and results are visualized in Figure 3.4 using ggplot2 (107).

4. Gut microbiome-wide search for bacterial azoreductases reveals inter-individual differences in prodrug delivery potential to the intestines

4.1 Background

Orally administered drugs are an attractive, non-invasive mode of delivery of pharmaceuticals to the intestines. The human gut microbiome plays an important role in drug metabolism, and is capable of activating (32,35), inactivating (111), and even toxifying (43) orally administered drugs. In fact, prodrugs containing azo bonds actually require azoreductive activity of bacteria to release the biologically active compound (32). For conditions such as Inflammatory Bowel Disease (IBD) and colorectal cancer (CRC), bacterial azoreductases have been exploited to deliver therapeutics such as 5-aminosalicylic acid (5-ASA), prednisolone (112), and celecoxib (113) topically to diseased interstitial tissues (**Figure 4.1**). A classic example of an azo-bonded prodrug is sulfasalazine, which consists of one sulfapyridine molecule bound to one 5-ASA molecule via an azo bond (**Figure 4.1A**). Following oral administration of sulfasalazine, bacterial azoreductases in the intestines reduce the azo bond, liberating 5-ASA and allowing it to confer its anti-inflammatory properties (114–116) topically on inflamed intestinal tissue (32). Direct oral administration of 5-ASA is not optimal, because the majority of the drug is absorbed in the small intestine and is sent through systemic circulation (117). More recent examples of azo-bonded prodrugs are OPN501 and celecoxib-5-ASA. OPN501 is made of up prednisolone, 5-ASA, and an inert cyclization product, DHQ. Upon azoreduction, 5-ASA is released and is

able to act upon the target tissue. Following a spontaneous cyclization reaction, prednisolone and DHQ are released where prednisolone can act upon the target tissue (112). Celecoxib-5-ASA is made up of celecoxib, 5-ASA, and DHQ, which are all released upon azoreduction and cyclization in a similar fashion to OPN501 (113).

A recent review article by Suzuki 2019 collected and curated information on experimentally confirmed azoreductases and described their preferred flavin cofactors and electron donors (118). They found that bacterial azoreductases cluster phylogenetically into four main clades, which they denote as clade I, clade II, clade III, and clade IV, while some enzymes form no clades. Further, each clade seems to harbor its own preference for either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as its flavin cofactor, and either NADH or NADPH as its preferred electron donor. Clades I, II and III are flavoproteins whereas Clade IV proteins are flavin-free. Clade I prefers NADPH as its electron donor, Clade II prefers NADH as its electron donor, and Clades III and IV generally use both. Of the 37 enzymes examined by Suzuki 2019, eight enzymes formed no distinct phylogenetic clade, and featured differences in their primary sequence length, flavin cofactor, and preferred electron donor. Additionally, many azoreductases participate in other reduction mechanisms including nitroreductase activity (119–126) and quinone oxidoreductase activity (127–132), suggesting that there are an array of reductase enzyme families participating in azo bond reduction in the human gut microbiome.

In addition to their relevance in drug delivery and efficacy, azoreductases have received attention from environmentalists and ecologists seeking solutions to azo dye pollution via bioremediation (133–135). Azo dyes such as Allura Red and Brilliant Black are commonly used in the food and textile industries, and waste from their production and usage

pollutes the environment (135,136). These and other dyes have received attention because of their potential to break down into toxic and carcinogenic metabolites (137,138) and because of their pollutant effects (139). This has led to a plethora of manuscripts characterizing the activity of novel azoreductases across the bacterial kingdom (124,140–152), many of which are explored in this work.

There is a growing body of literature suggesting that the composition of one's gut microbiota may have a significant influence on azo bond degradation in the gut. Sousa et al. 2014 (153) show significant intersubject variability in the rates of sulfasalazine metabolism in fecal slurries prepared from five different participants. More recently, Zahran et al. 2021 (154) showed a wide range in decolorization of the azo dye Brilliant Black among fecal slurries from 16 individuals; ranging from 0% decolorization to 100% decolorization. Even landmark studies demonstrating the efficacy of sulfasalazine treatment feature some individuals with little to no azo bond reduction (117). A non-trivial number of patients prescribed sulfasalazine (~10-45%) have reported adverse side effects such as headache, exanthema, and fever (155). These adverse reactions are linked with plasma levels of sulfapyridine (156–158), which is directly tied to the presence and activity of azoreducing bacteria in the gut. The presence or absence of azoreducing species in the human gut could contribute to these observed discrepancies in bioavailability of drugs (**Figure 4.1C**). By better understanding a patient's gut microbiome potential to reduce azo groups, clinicians could tune dosages of azo-bonded drugs to limit undesirable side effects of certain medications such as sulfasalazine.

It is important to know how prevalent azoreductase enzymes are in the human gut microbiome, and whether or not azoreduction is a core function of the human gut

microbiome. This has implications for the efficacy of existing prodrugs mentioned above, as well as for the development of future azo-bonded prodrugs. While azoreductases have been identified and characterized in various gut bacteria, the distribution of azoreductases has not been systematically explored across all known gut bacterial genomes. To address this gap, we conducted homologue search of known azoreductases across the unified human gastrointestinal genomes (UHGG) collection to identify putative azoreducing species in the human gut microbiome. We then assessed the relative abundance of these species in healthy, adenoma, CRC, and IBD populations. Finally, we examined the stability of known and putative azoreducing species in healthy, Crohn's disease (CD), and ulcerative colitis (UC) populations over time.

4.2 Results

4.2.1 Azoreductase genes are widely distributed in the human gut microbiome

To begin identifying putative azoreducing species of the gut microbiome, we searched the literature for experimentally verified azoreductase enzymes, collected amino acid sequences and metadata for these azoreductases, and clustered these sequences using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (159). The resulting sequence similarity network captured mechanistic preferences (flavin dependence and electron donor preference) of each azoreductase reported by Suzuki 2019 (118). We saw near complete concordance between the Clades I-IV and the sequence similarity clusters at 35% amino acid identity edge threshold, with the exception of Clade IV, which was split into two

separate clusters (**Figure 4.2**). Multiple sequence alignments were generated for each cluster shown in Figure 2 (with the exception of the "Other Azoreductases" group) using MUSCLE v3.8.425 (103). Profile Hidden Markov Models (pHMMs) were then trained on the multiple sequence alignments from the previous step using the HMMER (version 3.1b2) method hmmpress (91) and were queried against the Unified Human Gastrointestinal Genomes (UHGG) collection (90).

We searched for homologues of azoreductases across 4,644 representative genomes in the Unified Human Gastrointestinal Genomes (UHGG) collection (90) using pHMMs generated from sequences of experimentally validated azoreductase enzymes (**Figure 4.3**). This collection contains 204,938 genome sequences of bacteria known to inhabit the human gut, of which 4,644 are considered representative. For the remainder of this work, we refer to species receiving statistically significant (E-value < 1x10^-10) hits to azoreductase genes as "putative azoreducing species" or "putative azoreducing bacteria". Of the 4,644 genomes in the UHGG collection, there are 1,443 (31.1%) with one putative azoreductase gene, 343 (7.4%) with two or more putative azoreductases, and 372 (8.0%) with three or more putative azoreductases, indicating the potentially extensive ability of the gut microbiome to reduce azo bonds. Most notably, 364 genomes contain hits to the clade I profile, 452 contain hits to the clade II profile, 793 genomes contain hits to the clade III profile, 568 contain hits to the clade IVa profile, 410 contain hits to the clade IVb profile, 285 contain hits to the *mdaB* profile, and 477 contain hits to the *yieF* profile. Prominent phyla of the gut microbiome such as *Proteobacteria* and *Firmicutes* appear particularly rich with Clade I, Clade II, Clade III, Clade IVab, and FAD utilizing azoreductases (purple columns in **Figure 4.3**), suggesting that species in these phyla may encode multiple pathways of azoreduction (**Figure 4.3**).

4.2.2 Relative abundance of putative azoreducing bacteria is significantly higher than known azoreducing bacteria across healthy and disease populations

We next explored the relative abundance of putative and experimentally validated azoreducing species in 10,154 samples from 7,064 individuals from healthy, IBD, adenoma and CRC patient cohorts using publically available shotgun metagenomic datasets (59,60,93). This collection of data consists of samples from the Human Microbiome Project 2 (HMP2), Prospective Registry in IBD Study at MGH (PRISM), and curatedMetagenomicData. Putative azoreducing bacteria are considerably more abundant than known azoreducing bacteria in healthy, adenoma, and CRC individuals (all $p < 2.2x10^{-16}$, Mann-Whitney U test) (**Figure 4.4A**, all curves). We also found that putative azoreducing bacteria are significantly more abundant than known azoreducing species in individuals afflicted by Ulcerative Colitis and Crohn's disease. We saw this effect to varying degrees in samples obtained from the HMP2 (UC: $p < 6.0x10^{-5}$; CD: $p < 5.8x10^{-7}$) and PRISM (UC: p-value = $3.9x10^{-15}$, CD: p-value = $7.8x10^{-15}$) datasets.

Notable species identified through our homologue search include *Agathobacter rectale* (formerly known as *Eubacterium rectale*), *Bacteroides xylanisolvens* and *Fusobacterium nucleatum*, as these microbes are abundant across healthy, IBD, and CRC cohorts (**Figure 4.4B**). *Agathobacter rectale* has the highest mean relative abundance of any bacteria across all samples from cMD and PRISM. *Bacteroides xylanisolvens* has the highest mean relative abundance across HMP2 samples and second highest amongst cMD samples. *Fusobacterium nucleatum*, though it is lower in relative abundance than *A. rectale* and *B. xyanisolvens*, is known to be associated with colorectal cancer (79). These species are ideal candidates for experimental validation of azoreduction.

4.2.3 Relative abundance of putative azoreducing species fluctuates over time

We next sought to examine whether relative abundance levels of combined known and putative azoreducing species are stable or fluctuate over time. The HMP2 dataset provides a unique opportunity to examine the stability of individuals' gut microbiomes over time, as there are 18 individuals across healthy, UC and CD cohorts with at least 20 stool samples collected over a six month period. In order to examine the stability of azoreduction in the human gut, we compared the relative abundance of known and putative azoreducing species from these participants (**Figure 4.5**). The median relative abundance of combined azoreducing species ranges from 20.3 ± 3.58% to 33.9 ± 19.2% for nonIBD, 21.7 ± 7.6% to 49.0 ± 15.0% for UC, and 34.9 ± 6.39 to 62.3 ± 18.8 for CD subjects. Using linear mixed effects model ANOVA, we found that combined azoreducing species are significantly more abundant in CD subjects than in nonIBD subjects ($p = 0.002$) and are moderately more abundant in UC subjects than in nonIBD subjects ($p = 0.064$) (**Figure 4.5A**). Note that **Figure 4.5B** shows the same relative abundance values displayed in **Figure 4.5A** but over the course of the study, from collection 1 to collection 24.

4.3 Discussion

The presence of azoreducing bacteria in the human gut is necessary for effective delivery and activation of azobonded prodrugs. While estimates of bacterial species in the gut number in the thousands (58), only 54 bacterial species have been experimentally confirmed azoreducers. With our search for homologues of known azoreductases among a non-redundant set of 4,644 bacterial genomes, we identified an additional 1,958 putative azo reducing species (**Figure 4.3**). Additionally, our analysis of 10,154 samples from 7,064

individuals from healthy, IBD, adenoma and CRC patient cohorts across dozens of datasets showed that the relative abundance of putative azoreducing species far exceeds the relative abundance of known azoreducing species (**Figure 4.4**). However, we found that the presence of azoreducing bacteria appears unstable over time (**Figure 4.5**), which could certainly have an impact on the efficacy of azo bonded prodrugs such as sulfasalazine, OPN-501, and celecoxib-5-ASA.

Of the 1,958 putative azoreducing species identified in this work, we draw particular attention to *Agathobacter rectale*, *Bacteroides xylanisolvens*, and *Fusobacterium nucleatum* because of their high relative abundance levels across healthy and disease populations and the dearth of evidence surrounding their azoreductive abilities (**Figure 4.4**). In particular, the abundance of *Fusobacterium nucleatum* is positively correlated with colorectal cancer (62–64), present on the cancer itself, and possibly contributes to the etiology of the disease (65–67). Should *F. nucleatum* be experimentally validated as an azoreducing species, an anti-cancer prodrug could be designed to be specifically activated by *F. nucleatum* enabling specific tumor targeting. Further testing of these and potentially other putative azoreducers serve as future directions in the study of bacterial drug metabolism.

Importantly, we found that the relative abundance of known and putative azoreducing bacteria significantly higher in individuals with Crohn's Disease and are moderately higher in individuals with ulcerative colitis compared to healthy controls (**Figure 4.5**). This bodes well for the future of prodrug development since these therapies are intended to treat individuals afflicted with diseases affecting the intestines where azoreducing microbes are present to activate the prodrug. However, the cumulative relative abundance of known plus

putative azoreducing species does not appear particularly stable over time (**Figure 4.5B**) and future studies should explore whether there exists some minimum necessary abundance of azoreducing species for adequate prodrug metabolism and activation.

One limitation of this study is the lack of experimental evidence of azo bond reduction by putative azoreducing species identified in the homologue search step. This limitation can be addressed by screening highly abundant putative azoreducers such as *Agathobacter rectale* (formerly known as *Eubacterium rectale*). Azoreduction is a common feature in not only prodrug metabolism, but also in azo dye degradation, as many dyes commonly used in foods and textiles feature a nitrogen-nitrogen double bond (68). Thus, many of the azoreductase sequences used in our homologue search were experimentally verified by their ability to degrade azo dyes, and have not been tested on azo-bonded prodrugs. Bacterial azoreductases can have different affinities for different substrates (22,46,69–74), so, testing known azo dye reducing enzymes on prodrugs such as sulfasalazine would further refine our knowledge of bacterial prodrug activation.

4.5 Methods

4.5.1 Curation of pHMMs representing azoreductase enzymes

We searched the literature for known and experimentally validated species of bacteria which have azoreductase activity. Preliminary evidence for azoreductase gene sequence clustering is shown in Suzuki 2019 (11) where sequences were aligned and phylogenetically compared. Next, after collecting 40 sequences of experimentally validated azoreductases, we generated a sequence similarity network (SSN) using EFI-EST (55) at a 35% amino acid sequence identity threshold for identifying similar clusters of azoreductase genes. This

threshold corroborates the preliminary evidence for a diversity of azoreductase sequences put forth by Suzuki 2019 (11). With the exception of clade IVb sequences, which reached 31% sequence identity, all other clusters of genes had at least 35% sequence identity. The groups shown in **Figure 4.2** were pressed into profile Hidden Markov Models (pHMMs) using HMMER version 3.1b2 (57). Other genes that did not fall into the clade I - clade IV clusters (*arsH*, *yieF*, *mdaB, azo1*, *azoR*, etc.), were pressed into singular pHMMs and included in the homologue search.

4.5.2 Search for azoreductase genes across human gut microbial genomes

We searched pHMMs of known, experimentally validated azoreductases across 4,644 non-redundant genomes contained in the UHGG collection (58) using HMMER v3.1b2 (57). Alignments to queried pHMMs with *E*-value $<1x10^{-10}$ and 60% coverage of the query sequence were labeled as putative azoreductase gene sequences. Putative azoreducing bacterial species with experimentally confirmed azoreductase activity were then labeled as "known" azoreducing species, and classified separately from the putative azoreducing species. Putative azoreducing species across the bacterial taxonomy were visualized using the iTOL web interface (75) and prominent phyla of the gut microbiome were subsetted and presented in **Figure 4.3**.

4.5.3 Relative abundance estimation of gut bacteria

Raw sequencing reads for samples from curatedMetagenomicData (61), HMP2 (59) and PRISM (60) were downloaded and extracted with NCBI's SRA toolkit v2.10.9 (SRA-Tools, 2021). Quality control and adapter trimming of the fastq sequence files were done with the Trim Galore wrapper v0.6.6 (Babraham Bioinformatics-Trim Galore, 2021). To

remove potential human contaminants, quality trimmed reads were screened against the human genome (hg19) with Bowtie2 v2.4.2 (Langmead and Salzberg, 2012). Taxonomy profiling of the cleaned metagenomic reads were generated using Kraken 2 (2.0.8-beta) (Wood et al., 2019) to estimate the relative abundance of bacterial species present in each dataset. These relative abundances are then processed and plotted in Figure 4.4.

4.5.4 Statistical analysis of relative abundance data from cMD, HMP2 and PRISM datasets

Known (experimentally validated) azoreducing bacterial species were identified through extensive literature search and putative azoreducing species were identified through the homologue search step described above. Relative abundance values of known and putative azoreducing bacterial species were compared across healthy (nonIBD), colorectal cancer (CRC), adenoma, Ulcerative Colitis (UC), and Crohn's Disease (CD) subject populations using the Mann-Whitney U test (**Figure 4.4**). A non-parametric statistical test was chosen to compare relative abundances of known and putative azoreducing because the assumptions of independence and normally distributed data are violated with these datasets. Next, we compared the relative abundances of known and putative azoreducing species for HMP2 subjects with 20 or more stool samples taken over the course of the study. (**Figure 4.5**) A linear mixed effects model ANOVA was performed on this subset of HMP2 data to determine any statistically significant differences in relative abundance values across nonIBD, UC, and CD subjects. The R package lme4 (76) was used to fit the model to the data and the package lmerTest (77) was used to perform the ANOVA on the model. All statistical analyses were performed in R version 4.1.1 and are provided at https://github.com/dombraccia/Azoreductases.

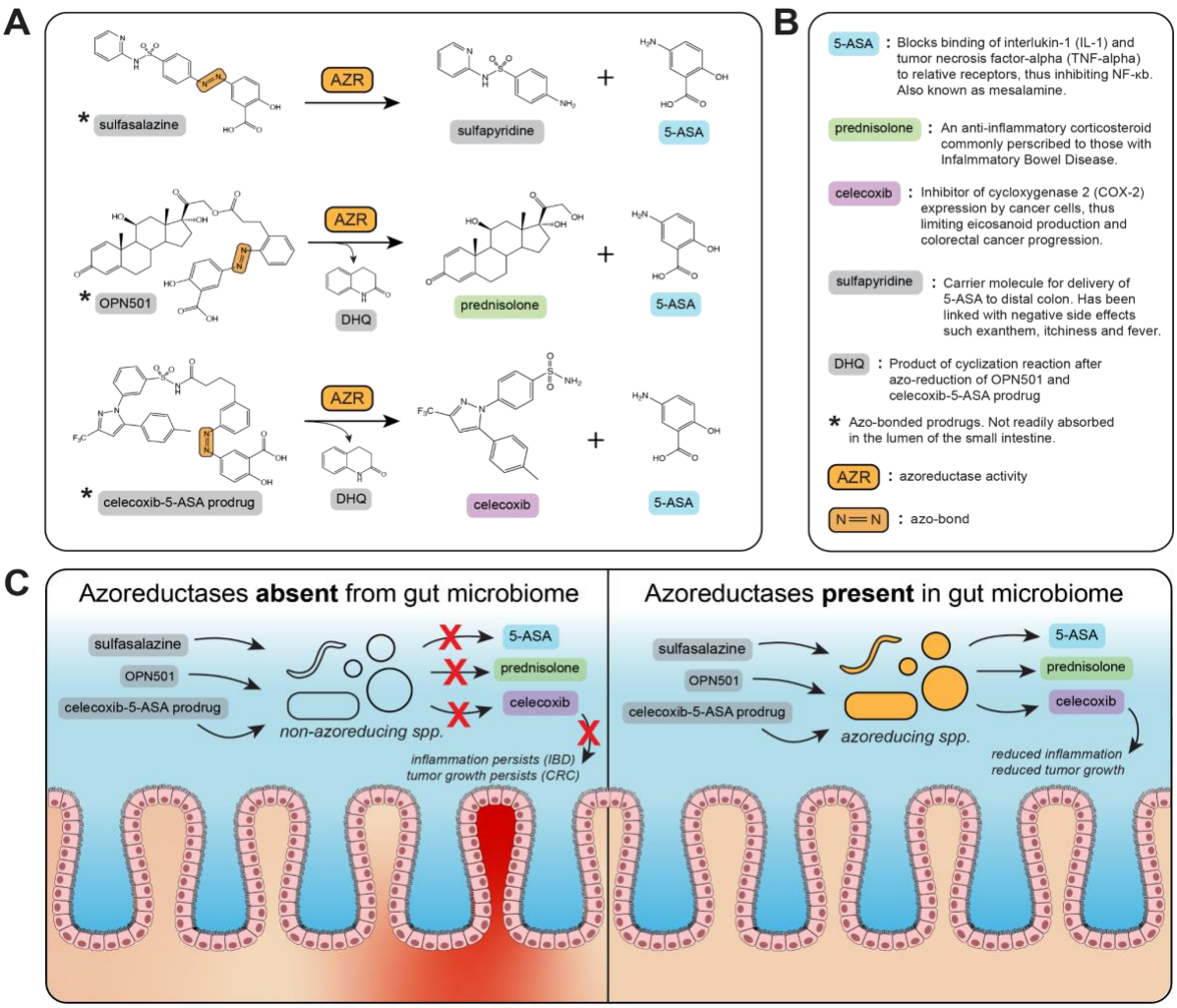

Figure 4.1. Mechanism of azoreduction by gut microbiota.

A) Pathways of azo-bonded drug activation by bacterial azoreductase activity. Dihydroquinolone (DHQ) is produced via cyclization of an intermediate of OPN501 and celecoxib-5-ASA prodrug following intramolecular lactamization (5,6). B) Description of downstream metabolites of bacterial azoreduction and the mechanisms of action in Inflammatory Bowel Disease (IBD) and colorectal cancer (CRC). References for each molecular function described in this subfigure: 5-ASA (78,79,9), prednisolone (80), celecoxib (81–83), sulfapyridine (84), DHQ (5,6). (C) Presence of azoreductase containing bacteria is required for prodrug activation in the IBD gut and CRC gut. Lack of azoreducing species renders orally administered prodrugs inactive and unabsorbed, leading to persistence of disease for IBD and CRC patients (left). Presence and activity of azoreducing species in the IBD and CRC gut leads to activation of orally ingested prodrugs, and site-specific delivery of active moieties to the colon (right).

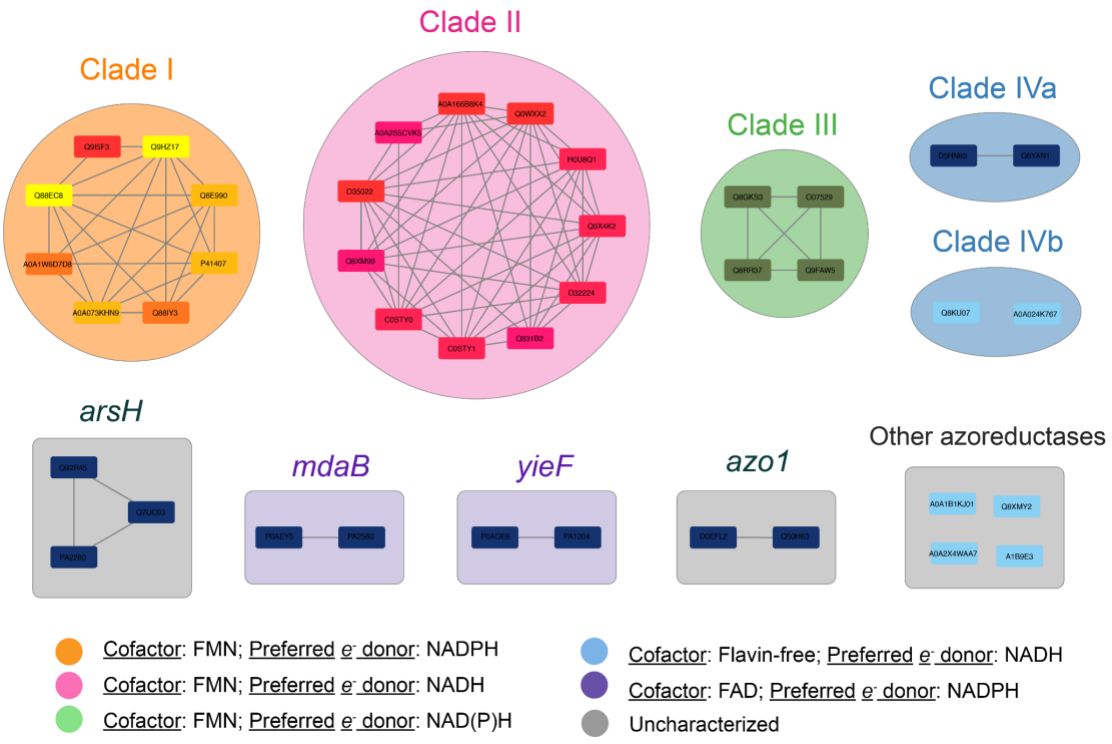

Figure 4.2. Bacterial azoreductases cluster by cofactor and electron donor preferences.

Following an extensive literature search for experimentally confirmed bacterial azoreductases, amino acid sequences of 40 azoreductase enzymes were collected and clustered with EFI-EST (159) at 35% sequence identity. Each node in the figure above is a single azoreductase gene and the edges between nodes indicate at least 35% sequence identity between the two amino acid sequences. The colored clusters, Clade I through Clade IVa&b, are groups of azoreductases previously described by Suzuki 2019 (118) as mechanistically similar groups based on cofactor and electron donor preferences. Clusters labeled with gene names (*mdaB*, *yieF*, etc.) represent homologous gene sequences found in two or more organisms. Each mechanistically characterized group of azoreductases were subsequently pressed into profile Hidden Markov Models (pHMMs) using HMMER version 3.1b2 (91) which formed the basis of the homologue search (**Figure S1)**. The group labeled "other azoreductases" contains sequences that did not fall into any cluster at the 35% identity threshold, and were pressed into singular pHMMs prior to the homologue search.

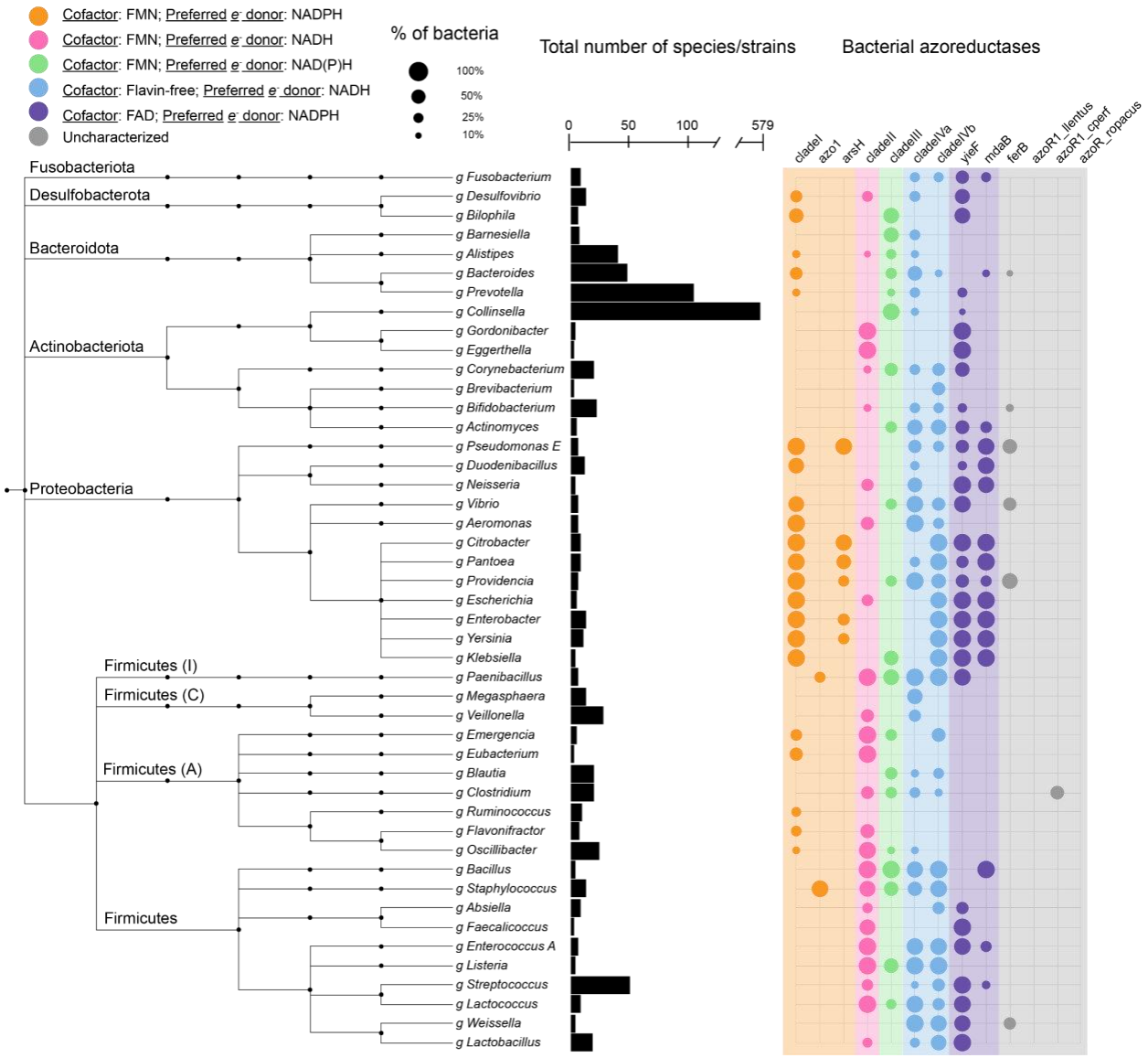

Figure 4.3. Azoreductases are widely distributed across gut bacterial taxonomy.

Presence/absence of azoreductases across prominent phyla of the human gut microbiome. The taxonomic tree is obtained from the Unified Human Gastrointestinal Genome collection (UHGG) (90) which is built on the Genome Taxonomy Database (GTDB) (110). Phyla names are annotated on the left side. Phyla names followed by a capital letter, e.g., Firmicutes (A), indicate a novel phyla classified by the GTDB-tk. The bar chart in the center indicates the number of species contained in each genus shown in the tree. The size of the circles indicates the number of species which contain hits to the azoreductase genes specified. The color of the circles indicates the cofactor and preferred electron (e^-) donor of the enzyme.

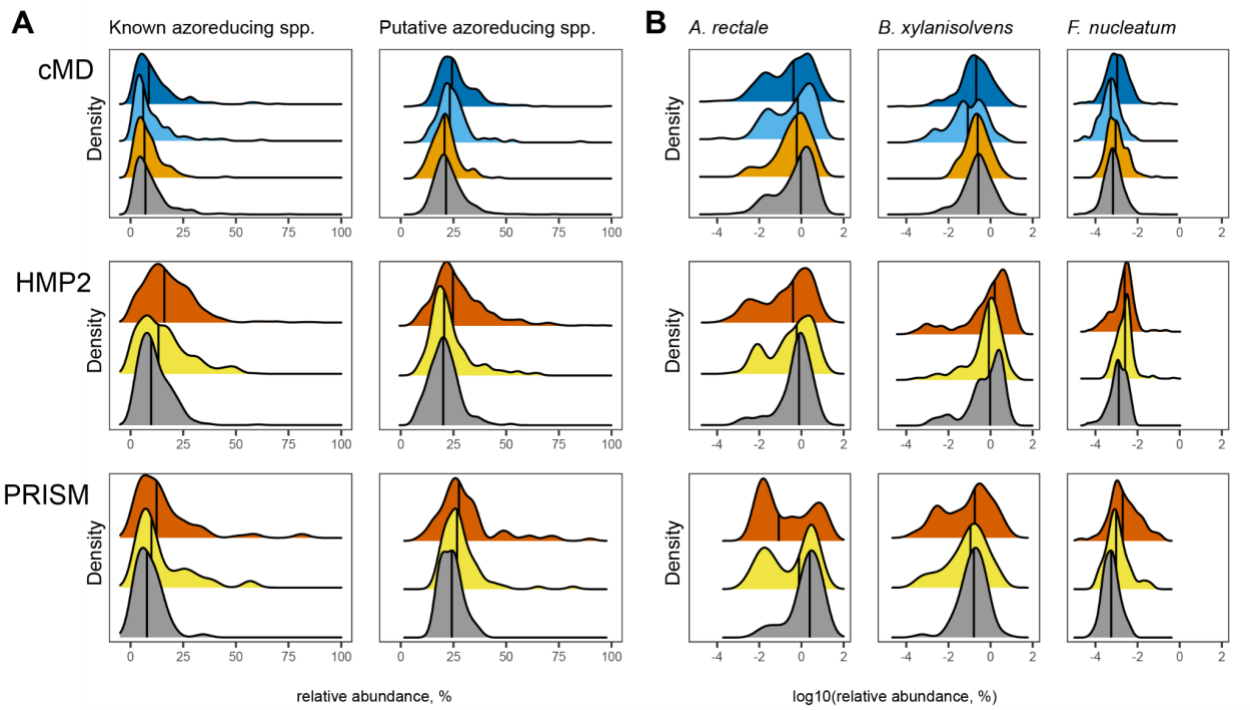

Figure 4.4. Relative abundance of putative azoreducers exceeds that of known azoreducers across healthy, IBD, adenoma and CRC populations.

Plots above show the distribution of relative abundance of known (experimentally verified) and putative azoreducing species. The y-axes display the smoothed kernel density estimated over the histogram of bacterial relative abundance values for each sample in the specified datasets (ggridges::stat_density_ridges method in R v4.1.1). Statistical significance asterisks ***** : $1x10^{-8} < p < 1x10^{-4}$; ****** : $1x10^{-15} < p < 1x10^{-14}$; ******* : $p < 2.2x10^{-16}$ (Wilcoxon rank sum test). Vertical lines indicate median values. A) Relative abundances of known azoreducing species (**Table S1**) and putative azoreducing species identified through homologue search. Metagenomic datasets explored: curatedMetagenomicData (cMD): 8,763 samples, 6,748 individuals (61); PRISM: 220 samples, 220 subjects (60); Human Microbiome Project 2 (HMP2): 1,112 samples, 104 subjects (59). B) Selected putative azoreducing species with the highest relative abundance. From left to right: *Agathobacter rectale*: highest mean relative abundance across samples from cMD and PRISM. *Bacteroides xylanisolvens*: highest mean relative abundance across HMP2 samples and second highest amongst cMD samples. *Fusobacterium nucleatum*: associated with colorectal cancer.

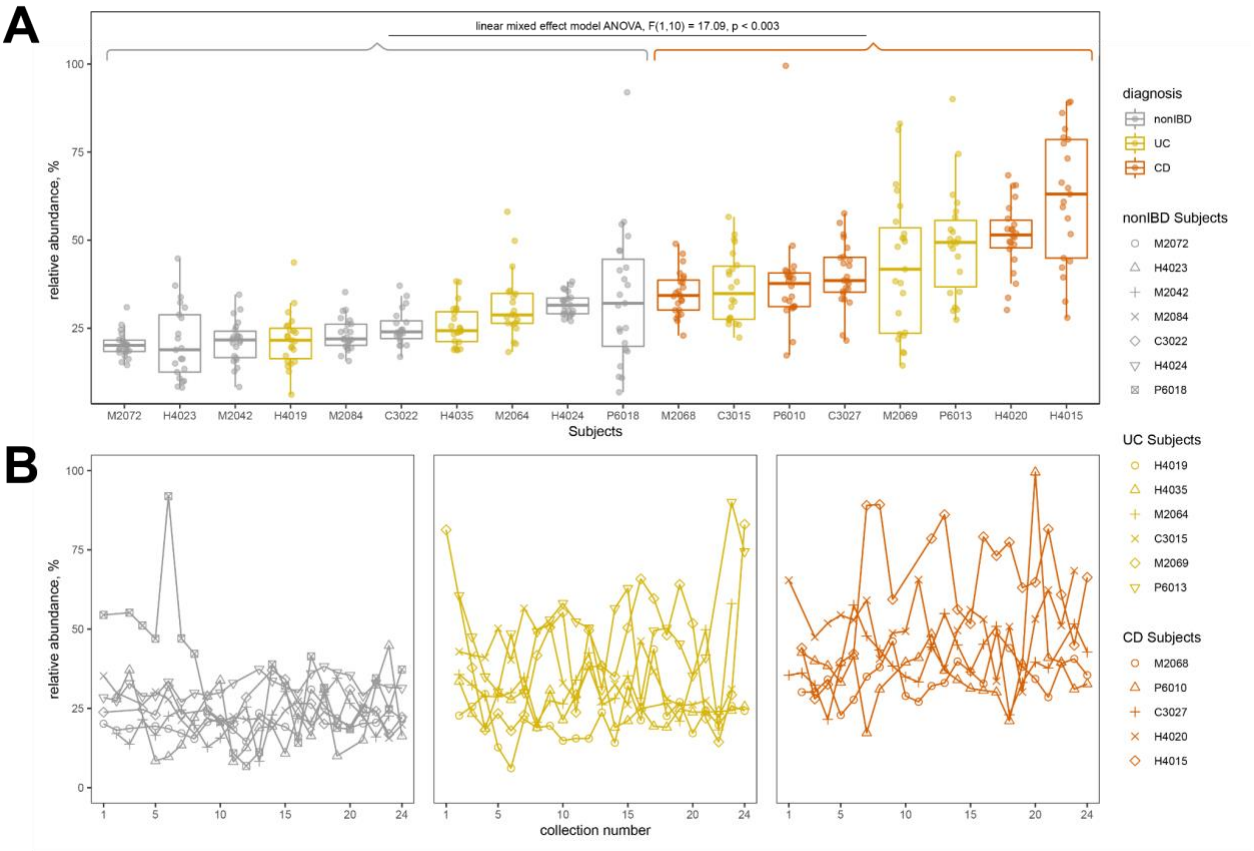

Figure 4.5. Known and putative azoreducing species are consistently more abundant in the IBD gut.

Relative abundance of known and putative azoreducing species in HMP2 subjects with more than 20 total stool collections. (A) Relative abundance of known and putative azoreducing species for qualifying participants across nonIBD, UC, and CD populations. Subjects with Crohn's Disease (CD) have significantly higher relative abundances of known+putative azoreducing species than healthy subjects (nonIBD) per linear mixed effects model ANOVA, $F(1,10) = 17.09$, $p < 0.003$. (B) Relative abundance of putative azoreducing species over time for healthy (nonIBD), Ulcerative Colitis (UC), and Crohn's Disease (CD) participants. Each line represents a single participant and each point is the summed relative abundance of known+putative azoreducing species at that collection point. The key on the right links relative abundance distributions for each subject with the same data point shown over collection numbers. Collections were taken approximately every 14 days.

5. Conclusion

In chapter 2, we perform patient stratification on 700 samples using untargeted metabolomics stools samples, resulting in patient labels for bilirubin reducers and nonreducers. These labels were used in conjunction with shotgun metagenomic sequencing data to perform differential species abundance analysis. We have prioritized strains higher in relative abundance in the bilirubin reducing group for follow-up experimental validation of bilirubin reduction. We have also simultaneously performed a differential gene abundance analysis using the same patient labels to identify candidate bilirubin reductase genes. In total, we have prioritized 20 strains and 18,893 genes for further testing. This work shows how multi-omic data can be used together to prioritize biological targets for experimental validation of health relevant functions.

Following the strain prioritization analysis outlined in section 2.2.2, we prioritized 20 bacterial species for experimental validation of bilirubin reduction. These strains will be grown anaerobically and tested for their bilirubin reduction ability. With known bilirubin reducing and nonreducing strains identified, we plan to perform a comparative genomics analysis between these sets of bacterial genomes to identify candidate bilirubin reductase genes. In a parallel computational workflow outlined in section 2.2.3, we identified 18,893 candidate bilirubin reductase genes. With the results of these two parallel prioritization pipelines in hand, we plan to further narrow down the set of candidate bilirubin reductase genes based on: (i) Enzyme Commission (EC) number relevancy, (ii) synteny within the bacterial genome, and (iii) containment in confirmed bilirubin reducing strains and exclusion from confirmed non-bilirubin reducing strains.

In chapter 3, we show that the relative abundance of primary putative cysteine-degrading bacteria is significantly higher than sulfate-reducing bacteria across healthy individuals as well as individuals with colorectal cancer and Inflammatory Bowel Disease. These results bolster previous studies suggesting the importance of dietary cysteine in gut microbial H_2S production. The systematic annotation of putative H_2S -producing species performed in this study can serve as a resource for future studies examining the links between H_2S and disease and could help these studies to tease out the concentration-dependent effects of H_2S on human health. Overall, this work informs future approaches to modulate gut microbial H_2S production via dietary interventions and may lead to an improved understanding of the complex interplay between H_2S and human health and disease.

In chapter 4, we show that azoreductases are widely distributed in the human gut microbiome and identify 1,958 putative azoreducing species from the 56 that have been experimentally confirmed prior to this work. We also show that 36.5% (715/1,958) of putative azoreducing species identified in this study contain hits to multiple azoreductases which suggests that many bacterial species are potentially able to metabolize a broad range of azo-bonded substrates. Furthermore, we show that the relative abundance of putative azoreducing species are present at equal or higher relative abundance levels in the CRC gut and IBD gut compared to healthy controls. However, we also show that the relative abundance of azoreducing species in the gut fluctuates over time, ranging from as low as 14.4% to as high as 84.0% in the same individual (UC, M2069). These large fluctuations of azoreducing bacteria in the gut could very well contribute to the variable efficacy observed for prodrugs such as sulfasalazine. Taken together, these results corroborate previous studies

suggesting that azoreduction is a core function of the human gut microbiome and highlight the possible utility of performing metagenomic sequencing on CRC and IBD patients prior to undergoing azo prodrug treatment.

List of Figures

Figure 2.1 Computational and statistical pipeline for patient stratification.

Figure 2.2 Computational pipeline for strain prioritization.

Figure 2.3. Results of strain prioritization for HMP2 and PRISM.

Figure 2.4. Custom gene catalogue generation pipeline.

Figure 2.5. Differential gene abundance analysis.

Figure 3.1. H_2S production via cysteine degradation in the human gut microbiome.

Figure 3.2. Presence of cysteine-degrading genes across the human gut microbiome.

Figure 3.3. Putative primary and secondary cysteine-degrading bacteria are more prevalent than (SRB) among individuals with IBD, CRC, and healthy controls.

Figure 3.4. Cysteine-degrading genes are actively expressed in healthy adults.

Figure 3.5. Computational workflow.

Figure 4.1. Mechanism of azoreduction by gut microbiota.

Figure 4.2. Bacterial azoreductases cluster by cofactor and electron donor preferences.

Figure 4.3. Azoreductases are widely distributed across gut bacterial taxonomy.

Figure 4.4. Relative abundance of putative azoreducers exceeds that of known azoreducers across healthy, IBD, adenoma and CRC populations.

Figure 4.5 Known and putative azoreducing species are consistently more abundant in the IBD gut.

List of Abbreviations

5-aminosalicylic acid (5-ASA)

Bilirubin reducing (BR)

Bilirubin Reductase Candidate 4 (BRC4)

colorectal cancer (CRC)

Crohn's disease (CD)

curatedMetagenomicData (cMD)

Cysteine desulfurase activity (CYD)

dissimilatory sulfate reduction (DSR)

Enzyme Commission (EC)

Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST)

Flavin adenine dinucleotide (FAD)

Flavin mononucleotide (FMN)

Genome Taxonomy Database (GTDB)

Genome Taxonomy Database Tool Kit (GTDB-tk)

Hidden Markov Models (HMMs)

Human Microbiome Project 2 (HMP2)

Hydrogen (H_2)

Carbon dioxide (CO_2)

Hydrogen sulfide (H_2S)

Inflammatory Bowel Disease (IBD)

N-acetyl-p-benzoquinoneimine (NAPQI)

Nicotinamide adenine dinucleotide (NADH)

Nicotinamide adenine dinucleotide phosphate (NADPH)

Non-bilirubin reducing (nonBR)

non-IBD participant, healthy (nonIBD)

profile Hidden Markov Models (pHMMs)

Prospective Registry in IBD Study at MGH (PRISM)

pyridoxal-5′-phosphate (PLP)

Red blood cells (RBCs)

Sequence Alignment Map (SAM)

Sequence Read Archive (SRA)

Sulfate-reducing bacteria (SRB)

The Health Professionals Follow-up Study (HPFS)

Transcripts per million (TPM)

Trimethylamine (TMA)

UDP-glucuronosyltransferase 1 (UGT1A1)

Ulcerative colitis (UC)

Unified Human Gastrointestinal Genomes (UHGG)

Appendices

Appendix 1: Characterization of H_2S producing genes included in the search space

For each gene considered, we stated 1. the main species the gene is found in (limited to gut inhabiting bacteria), 2. whether the downstream enzyme is dependent on pyridoxal 5'-phosphate (PLP) binding, and 3. the primary function of the gene with references.

- **Green** highlighted genes indicate that H_2S output is central to the **primary** function of the gene and/or these genes are upregulated in the presence of cysteine.
- **Yellow** highlighted genes indicate that H_2S output is a **secondary** function of the gene and/or is closely related to cysteine degradation. Most of these genes, apart from *malY*, are involved in some component of the transulfuration pathway and additionally exhibit cysteine desulfurase activity.
- **Red** highlighted genes indicate that the gene has been shown to produce H_2S from cysteine, but the primary function of the gene is unrelated to H_2S production and/or the gene does appear to be upregulated in the presence of cysteine. We refer to these genes as **erroneous** H2S producing genes.

gene abbreviation	**Full gene name**

dcyD	**Cysteine desulfhydrase**

- Main species:
  - *Klebsiella pneumoniae (many strains)*
  - *Escherichia coli (K-12)*
- UniProt Entry: https://www.uniprot.org/uniprot/A6TB69
- PLP-dependent: **YES**
- Primary function: catalyzes the the beta elimination of D-cysteine and other D-cysteine derivatives. Potentially useful for detoxification of D-cysteine. (160,161)

***yhaOM* operon**	**Unnamed proteins**

- Main species:
  - *Escherichia coli (K-12)*
- PLP-dependent: **NO**
- Primary function: yhaO and yhaM are genes in the yhaOM operon, regulated by decR in E. coli (102). *yhaO* is likely a serine/cysteine transporter protein

and *yhaM* has cysteine desulfurase activity (102). Of all the cysteine desulfurase active proteins in *E. coli*, *yhaM* has been proposed to be the primary cysteine desulfurase protein (162), as many of the other genes identified by Awona et al. (99,163) have other primary functions such as Fe-S complex formation in the case of *iscS*.

mgl	**Methionine gamma-lyase**

- Main species:
  - *Fusobacterium nucleatum*
  - *Porphyromonas gingivalis*
  - *Treponema denticola*
- PLP-dependent: **YES**
- Primary function: The above species are the most frequently encountered species in the human oral microbiome, and have been implicated in periodontitis (101). Due to the mucosal-degrading activity of methanethiol (CH_3SH) and hydrogen sulfide (H_2S), it is believed by some that *mgl* is actively expressed in the presence of methionine and cysteine and acts as a pathogenicity factor (164).

aspC* + *sseA	**Cysteine aminotransferase + 3-mercaptopyruvate sulfurtransferase**

- Main species:
  - *Escherichia coli (K-12)*
  - *aspC* found in *E. coli K-12*
  - Many other species have *sseA* predicted based on homology including: *Yersinia spp., Salmonella spp., Brucella spp., Clostridium spp.* and *Klebsiella spp. (100)*
- PLP-dependent: *aspC* – **YES**; *sseA* – **NO**
- Primary function: Proposed by Shatalin et al. (100), H_2S production by *sseA* leads to decreased antibiotic sensitivity. However, it appears that other modes of H_2S production are more prominent in anerobic environments (165). This mechanism is more complicated, because *sseA* acts on 3-mercaptopyruvate, which is produced from cysteine by a cysteine aminotransferase, such as *aspC* (166).

metC	**Cystathionine beta-lyase**

- Main species:
  - *Escherichia coli (K-12)*
  - *Bacillus subtilis (168)*
- PLP-dependent: **YES**

- Primary function: Involved in the transulfuration pathway (cysteine to methionine interconversion). Specifically, *metC* reversibly converts cystathionine to homocysteine (108,163,167). Any cysteine desulfhydrase activity resulting in H2S production seems to be secondary to the primary function of cystathionine to homocysteine interconversion (102).

malY	**Cystathionine beta-lyase like; repressor of maltose regulon**

- Main species:
  - *Escherichia coli (K-12)*
- PLP-dependent: **YES**
- Primary function: *malY* is one of the few instances of a truly bi-functional protein on this list, serving as both a cystathionine beta-lyase (*metC*-like) and as a repressor of the maltose regulon responsible for the uptake and metabolism of maltose and maltodextrins (168,169). *malY* can recapitulate cystathionine beta-lyase activity in *metC* mutant strains of *E. coli*.

cysK	**Cysteine synthase A**

- Main species:
  - *Escherichia coli (K-12)*
  - *Bacillus subtilis (168)*
- PLP-dependent: **YES**
- Primary function: converts O-acetyl-serine + H_2S -> acetate + cysteine (108), however, the pathway appears to be reversible, resulting in H2S production from cysteine (99,102). Additionally, appears to be crucial in contact dependent growth inhibition of competing bacteria (170).

cysM	**Cysteine synthase B**

- Main species:
  - *Escherichia coli (K-12)*
  - *Bacillus subtilis (168)*
- PLP-dependent: **YES**
- Primary function: converts O-acetyl-serine + H_2S -> acetate + cysteine (108), however, the pathway appears to be reversible, resulting in H2S production from cysteine (99,102).

mccB	**Cystathionine gamma-lyase**

- Main species:
  - *Bacillus subtilis (168)*
- PLP-dependent: **YES**

- Primary function: Catalyzes the conversion of cystathionine -> cysteine. Can also catalyze a similar gamma-lyase reaction converting homocysteine + H_2O -> H_2S + 2-oxobutanoate + ammonia + pyruvate (108). Due to the potentially bifunctionality of this gene, it may be upgraded from erroneous to secondary production of H_2S.

tnaA	**Tryptophanase**

- Main species:
  - *Escherichia coli (K-12)*
  - *Proteus vulgaris*
  - *Shigella flexneri*
  - *Klebsiella aerogenes*
- PLP-dependent: **YES**
- Primary function: Catalyzes the degradation of H_2O + tryptophan -> indole + ammonia + pyruvate (171). *tnaA* has been shown to perform cysteine dehydratase activity under certain circumstances (163), but the upregulation of *tnaA* expression and activity could not be replicated in future studies (162). It is possible that the primary function of *tnaA* is tryptophan degradation and any cysteine desulhydrase activity is erroneous.

iscS	**Cysteine desulfurase**

- Main species:
  - *Escherichia coli (K-12)*
  - (very common gene in all of life)
- PLP-dependent: **YES**
- Primary function: *iscS* is potentially the most clear-cut case of erroneous H2S production by gut bacteria. *iscS* is a master enzyme that transfers sulfur from cysteine to aid in Fe-S cluster assembly, a key structure in the formation of many proteins (172) (and many other references). However, it should be noted that the production of H_2S mediated by *iscS* in E. coli can still prove to be quite substantial in anaerobic and high cysteine conditions (165).

mccA	**O-acetylserine dependent cystathionine beta-synthase**

- Main species:
  - *Bacillus subtilis (168)*
  - *Staphylococcus aureus*
- PLP-dependent: **YES**
- Primary function: Converts O-acetylserine and homocysteine to cystathionine (108). Given the primary function of the enzyme, it appears that any cysteine desulfurase activity is either secondary or erroneous.

Bibliography

1. Zhu B, Wang X, Li L. Human gut microbiome: the second genome of human body. Protein Cell. 2010 Aug;1(8):718–25.

2. Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, et al. The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. Cell Host Microbe. 2019 Oct 9;26(4):527-541.e5.

3. Hallen-Adams HE, Suhr MJ. Fungi in the healthy human gastrointestinal tract. Virulence. 2017 Apr 3;8(3):352–8.

4. Heintz-Buschart A, Wilmes P. Human Gut Microbiome: Function Matters. Trends Microbiol. 2018 Jul 1;26(7):563–74.

5. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804–10.

6. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, et al. The Human Microbiome Project: a community resource for the healthy human microbiome. 2012;

7. Tierney BT, Yang Z, Luber JM, Beaudin M, Wibowo MC, Baek C, et al. The Landscape of Genetic Content in the Gut and Oral Human Microbiome. Cell Host Microbe. 2019 Aug 14;26(2):283-295.e8.

8. Matsuoka K, Kanai T. The gut microbiota and inflammatory bowel disease. Semin Immunopathol. 2015 Jan 1;37(1):47–55.

9. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota disbiosis is associated with colorectal cancer. Front Microbiol [Internet]. 2015 [cited 2022 Jun 18];6. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2015.00020

10. Vogt NM, Kerby RL, Dill-McFarland KA, Harding SJ, Merluzzi AP, Johnson SC, et al. Gut microbiome alterations in Alzheimer's disease. Sci Rep. 2017;7(1):1–11.

11. Chok KC, Ng KY, Koh RY, Chye SM. Role of the gut microbiome in Alzheimer's disease. Rev Neurosci. 2021;32(7):767–89.

12. Khan I, Ullah N, Zha L, Bai Y, Khan A, Zhao T, et al. Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD Treatment Targeting the Gut Microbiome. Pathogens. 2019 Sep;8(3):126.

13. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol. 2019 Nov;16(11):690–704.

14. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016 Dec;8(1):1–11.

15. Fu X, Liu Z, Zhu C, Mou H, Kong Q. Nondigestible carbohydrates, butyrate, and butyrate-producing bacteria. Crit Rev Food Sci Nutr. 2019 Jun 27;59(sup1):S130–52.

16. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin Biotechnol. 2013 Apr;24(2):160–8.

17. Zhao Q, Elson CO. Adaptive immune education by gut microbiota antigens. Immunology. 2018;154(1):28–37.

18. Lathrop SK, Bloom SM, Rao SM, Nutsch K, Lio CW, Santacruz N, et al. Peripheral education of the immune system by colonic commensal microbiota. Nature. 2011;478(7368):250–4.

19. Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev. 2019 Aug 21;83(3):e00007-19.

20. Romano KA, Vivas EI, Amador-Noguez D, Rey FE. Intestinal Microbiota Composition Modulates Choline Bioavailability from Diet and Accumulation of the Proatherogenic Metabolite Trimethylamine-N-Oxide. mBio. 2015 Mar 17;6(2):e02481-14.

21. Haiser HJ, Seim KL, Balskus EP, Turnbaugh PJ. Mechanistic insight into digoxin inactivation by Eggerthella lenta augments our understanding of its pharmacokinetics. Gut Microbes. 2014 Mar 1;5(2):233–8.

22. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, et al. Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr. 2018 Feb 1;57(1):1–24.

23. Danilova NA, Александровна ДН, Abdulkhakov SR, Рустамович АС, Grigoryeva TV, Владимировна ГТ, et al. Markers of dysbiosis in patients with ulcerative colitis and Crohn's disease. Ter Arkh. 2019 Apr 15;91(4):13–20.

24. Roberts MS, Magnusson BM, Burczynski FJ, Weiss M. Enterohepatic Circulation. Clin Pharmacokinet. 2002 Aug 1;41(10):751–90.

25. Fevery J. Bilirubin in clinical practice: a review. Liver Int. 2008;28(5):592–605.

26. Fahmy K, Gray CH, Nicholson DC. The reduction of bile pigments by faecal and intestinal bacteria. Biochim Biophys Acta BBA - Gen Subj. 1972 Mar 30;264(1):85–97.

27. Suarez F, Furne J, Springfield J, Levitt M. Insights into human colonic physiology obtained from the study of flatus composition. Am J Physiol-Gastrointest Liver Physiol. 1997 May 1;272(5):G1028–33.

28. Levitt MD, Bond JH. Volume, Composition, and Source of Intestinal Gas. Gastroenterology. 1970 Dec;59(6):921–9.

29. Nava GM, Carbonero F, Croix JA, Greenberg E, Gaskins HR. Abundance and diversity of mucosa-associated hydrogenotrophic microbes in the healthy human colon. ISME J. 2012 Jan;6(1):57–70.

30. Ijssennagger N, van der Meer R, van Mil SWC. Sulfide as a Mucus Barrier-Breaker in Inflammatory Bowel Disease? Trends Mol Med. 2016 Mar 1;22(3):190–9.

31. Spanogiannopoulos P, Bess EN, Carmody RN, Turnbaugh PJ. The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. Nat Rev Microbiol. 2016 Apr;14(5):273–87.

32. Peppercorn MA, Goldman P. The Role of Intestinal Bacteria in the Metabolism of Salicylazosulfapyridine. J Pharmacol Exp Ther. 1972 Jun 1;181(3):555–62.

33. Gingell R, Bridges JW. Intestinal Azo-reduction and Glucuronide Conjugation of Prontosil. Xenobiotica. 1973 Jan;3(9):599–604.

34. Klotz U, Maier K, Fischer C, Heinkel K. Therapeutic Efficacy of Sulfasalazine and Its Metabolites in Patients with Ulcerative Colitis and Crohn's Disease. N Engl J Med. 1980 Dec 25;303(26):1499–502.

35. Colebrook L, Buttle GAH, O'Meara RAQ. THE MODE OF ACTION OF p-AMINOBENZENESULPHONAMIDE AND PRONTOSIL IN HÆMOLYTIC STREPTOCOCCAL INFECTIONS. The Lancet. 1936 Dec 5;228(5910):1323–6.

36. Fuller AT. IS p-AMINOBENZENESULPHONAMIDE THE ACTIVE AGENT IN PRONTOSIL THERAPY ? The Lancet. 1937 Jan;229(5917):194–8.

37. Ehle M, Patel C, Giugliano RP. Digoxin: Clinical Highlights: A Review of Digoxin and Its Use in Contemporary Medicine. Crit Pathw Cardiol. 2011 Jun;10(2):93–8.

38. Dobkin JF, Saha JR, Butler VP, Neu HC, Lindenbaum J. Digoxin-Inactivating Bacteria: Identification in Human Gut Flora. Science. 1983 Apr 15;220(4594):325–7.

39. Geographic differences in digoxin inactivation, a metabolic activity of the human anaerobic gut flora. | Gut [Internet]. [cited 2022 Jun 20]. Available from: https://gut-bmj-com.proxy-um.researchport.umd.edu/content/30/7/971.abstract

40. Rowland IR. Factors affecting metabolic activity of the intestinal microflora. Drug Metab Rev. 1988;19(3–4):243–61.

41. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Science. 2013;341(6143):295–8.

42. Tanizawa A, Fujimori A, Fujimori Y, Pommier Y. Comparison of topoisomerase I inhibition, DNA damage, and cytotoxicity of camptothecin derivatives presently in clinical trials. JNCI J Natl Cancer Inst. 1994;86(11):836–42.

43. Wallace BD, Wang H, Lane KT, Scott JE, Orans J, Koo JS, et al. Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. Science [Internet]. 2010 Nov 5 [cited 2021 Sep 3]; Available from: https://www.science.org/doi/abs/10.1126/science.1191175

44. Ghanem CI, Pérez María J, Manautou JE, Mottino AD. ACETAMINOPHEN; FROM LIVER TO BRAIN: NEW INSIGHTS INTO DRUG PHARMACOLOGICAL ACTION AND TOXICITY. Pharmacol Res. 2016 Jul;109:119–31.

45. Jollow DJ, Mitchell JR, Potter WZ, Davis DC, Gillette JR, Brodie BB. Acetaminophen-Induced Hepatic Necrosis. Ii. Role of Covalent Binding in Vivo. J Pharmacol Exp Ther. 1973 Oct 1;187(1):195–202.

46. Kruse A, Uehlinger DE, Gotch F, Kotanko P, Levin NW. Red Blood Cell Lifespan, Erythropoiesis and Hemoglobin Control. Hemodial - Basic Res Clin Trials. 2008;161:247–54.

47. Wang X, Chowdhury JR, Chowdhury NR. Bilirubin metabolism: Applied physiology. Curr Paediatr. 2006 Feb 1;16(1):70–4.

48. Brink MA, Slors JFM, Keulemans YCA, Mok KS, De Waart DR, Carey MC, et al. Enterohepatic cycling of bilirubin: A putative mechanism for pigment gallstone formation in ileal Crohn's disease. Gastroenterology. 1999 Jun 1;116(6):1420–7.

49. Fargo MV, Grogan SP, Saguil A. Evaluation of Jaundice in Adults. Am Fam Physician. 2017 Feb 1;95(3):164–8.

50. Dennery PA, Seidman DS, Stevenson DK. Neonatal Hyperbilirubinemia. N Engl J Med. 2001 Feb 22;344(8):581–90.

51. Lapidus A, Bångstad M, Åström M, Muhrbeck O. The prevalence of gallstone disease in a defined cohort of patients with Crohn's disease. Am J Gastroenterol. 1999 May 1;94(5):1261–6.

52. Harb T, Matsuyama M, David M, Hill RJ. Infant Colic—What works: A Systematic Review of Interventions for Breast-fed Infants. J Pediatr Gastroenterol Nutr. 2016 May;62(5):668–86.

53. Sung V, D'Amico F, Cabana MD, Chau K, Koren G, Savino F, et al. Lactobacillus reuteri to Treat Infant Colic: A Meta-analysis. Pediatrics. 2018 Jan 1;141(1):e20171811.

54. Szajewska H, Canani RB, Guarino A, Hojsak I, Indrio F, Kolacek S, et al. Probiotics for the prevention of antibiotic-associated diarrhea in children. J Pediatr Gastroenterol Nutr. 2016;62(3):495–506.

55. Guo Q, Goldenberg JZ, Humphrey C, El Dib R, Johnston BC. Probiotics for the prevention of pediatric antibiotic-associated diarrhea. Cochrane Database Syst Rev. 2019;(4).

56. van den Akker CH, Van Goudoever JB, Szajewska H, Embleton ND, Hojsak I, Reid D, et al. Probiotics for preterm infants: a strain-specific systematic review and network meta-analysis. J Pediatr Gastroenterol Nutr. 2018;67(1):103–22.

57. Morgan RL, Preidis GA, Kashyap PC, Weizman AV, Sadeghirad B, Chang Y, et al. Probiotics reduce mortality and morbidity in preterm, low-birth-weight infants: a systematic review and network meta-analysis of randomized trials. Gastroenterology. 2020;159(2):467–80.

58. Sharif SP, Friedmacher F, Amin A, Zaki RA, Hird MF, Khashu M, et al. Low serum albumin concentration predicts the need for surgical intervention in neonates with necrotizing enterocolitis. J Pediatr Surg. 2020;55(12):2625–9.

59. Proctor L, Huttenhower C. The Integrative Human Microbiome Project. Nature. 2019 May;569(7758):641–8.

60. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019 Feb;4(2):293–305.

61. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015 Oct;12(10):902–3.

62. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLOS Comput Biol. 2021 Nov 16;17(11):e1009442.

63. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019 Dec;20(1):257.

64. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674–6.

65. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

66. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr;14(4):417–9.

67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550.

68. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018 Sep 1;34(17):i884–90.

69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9.

70. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

71. McKnight PE, Najab J. Mann-Whitney U Test. In: The Corsini Encyclopedia of Psychology [Internet]. John Wiley & Sons, Ltd; 2010 [cited 2022 Jul 5]. p. 1–1. Available from: http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470479216.corpsy0524

72. Wang R. Hydrogen Sulfide: The Third Gasotransmitter in Biology and Medicine. Antioxid Redox Signal. 2009 Oct 21;12(9):1061–4.

73. Kimura H. Hydrogen Sulfide: From Brain to Gut. Antioxid Redox Signal. 2009 Oct 5;12(9):1111–23.

74. Percudani R, Peracchi A. A genomic overview of pyridoxal-phosphate-dependent enzymes. EMBO Rep. 2003 Sep 1;4(9):850–4.

75. Wang R. Physiological Implications of Hydrogen Sulfide: A Whiff Exploration That Blossomed. Physiol Rev. 2012 Apr;92(2):791–896.

76. Shekarriz R, Friedrichsen DM, Brooks B, Silaski G, Rios L, Wiest E, et al. Sensor of transdermal biomarkers for blood perfusion monitoring. Sens Bio-Sens Res. 2020 Jun 1;28:100328.

77. Furne J, Springfield J, Koenig T, DeMaster E, Levitt MD. Oxidation of hydrogen sulfide and methanethiol to thiosulfate by rat tissues: a specialized function of the colonic mucosa. Biochem Pharmacol. 2001 Jul;62(2):255–9.

78. Blachier F, Andriamihaja M, Larraufie P, Ahn E, Lan A, Kim E. Production of hydrogen sulfide by the intestinal microbiota and epithelial cells and consequences for the colonic and rectal mucosa. Am J Physiol-Gastrointest Liver Physiol. 2021 Feb 1;320(2):G125–35.

79. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012 Feb 1;22(2):299–306.

80. Carbonero F, Benefiel AC, Alizadeh-Ghamsari AH, Gaskins HR. Microbial pathways in colonic sulfur metabolism and links with health and disease. Front Physiol [Internet]. 2012 Nov 28 [cited 2020 Aug 13];3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3508456/

81. Gibson G r., Macfarlane G t., Cummings J h. Occurrence of sulphate-reducing bacteria in human faeces and the relationship of dissimilatory sulphate reduction to methanogenesis in the large gut. J Appl Bacteriol. 1988 Aug 1;65(2):103–11.

82. Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, et al. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. Int J Syst Evol Microbiol. 70(11):5972–6016.

83. Gibson GR, Cummings JH, Macfarlane GT. Competition for hydrogen between sulphate-reducing bacteria and methanogenic bacteria from the human large intestine. J Appl Bacteriol. 1988 Sep;65(3):241–7.

84. Deng YF, Liao XD, Wang Y, Liang JB, Tufarelli V. Prebiotics Mitigate In Vitro Sulfur-Containing Odour Generation in Caecal Content of Pigs. Ital J Anim Sci. 2015 Jan 1;14(1):3762.

85. Yao CK, Rotbart A, Ou JZ, Kalantar-Zadeh K, Muir JG, Gibson PR. Modulation of colonic hydrogen sulfide production by diet and mesalazine utilizing a novel gas-profiling technology. Gut Microbes. 2018 May 9;9(6):510–22.

86. Miller TL, Wolin MJ, Macario EC de, Macario AJ. Isolation of Methanobrevibacter smithii from human feces. Appl Environ Microbiol. 1982 Jan 1;43(1):227–32.

87. Lin R, Liu W, Piao M, Zhu H. A review of the relationship between the gut microbiota and amino acid metabolism. Amino Acids. 2017 Dec 1;49(12):2083–90.

88. Silvester KR, Cummings JH. Does digestibility of meat protein help explain large bowel cancer risk? Nutr Cancer. 1995 Jan 1;24(3):279–88.

89. Levine J, Ellis CJ, Furne JK, Springfield J, Levitt MD. Fecal Hydrogen Sulfide Production in Ulcerative Colitis. Am J Gastroenterol. 1998 Jan;93(1):83–7.

90. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol. 2020 Jul 20;1–10.

91. HMMER [Internet]. [cited 2021 Aug 18]. Available from: http://hmmer.org/

92. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe. 2015 Oct 14;18(4):489–500.

93. Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, et al. Accessible, curated metagenomic data through ExperimentHub. Nat Methods. 2017 Nov;14(11):1023–4.

94. Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory bowel disease. Nat Rev Microbiol. 2019 Aug;17(8):497–511.

95. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol. 2018 Mar;3(3):356–66.

96. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014 Jan 23;505(7484):559–63.

97. Delignette-Muller ML, Dutang C. fitdistrplus: An R Package for Fitting Distributions. J Stat Softw. 2015 Mar 20;64(1):1–34.

98. Nguyen LH, Abu-Ali G, Mehta RS, Lloyd-Price J, Song M, Yan Y, et al. Su1940 - Dietary Patterns, Sulfur Intake, and the Abundance of Sulfate-Reducing Bacteria. Gastroenterology. 2018 May;154(6):S-640.

99. Awano N, Wada M, Mori H, Nakamori S, Takagi H. Identification and Functional Analysis of Escherichia coli Cysteine Desulfhydrases. Appl Environ Microbiol. 2005 Jul 1;71(7):4149–52.

100. Shatalin K, Shatalina E, Mironov A, Nudler E. H2S: A Universal Defense Against Antibiotics in Bacteria. Science. 2011 Nov 18;334(6058):986–90.

101. Suwabe K, Yoshida Y, Nagano K, Yoshimura F. Identification of an l-methionine γ-lyase involved in the production of hydrogen sulfide from l-cysteine

in Fusobacterium nucleatum subsp. nucleatum ATCC 25586. Microbiology,. 2011;157(10):2992–3000.

102. Shimada T, Tanaka K, Ishihama A. Transcription factor DecR (YbaO) controls detoxification of L-cysteine in Escherichia coli. Microbiology. 2016 Sep 1;162(9):1698–707.

103. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 1;32(5):1792–7.

104. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019 Jul 2;47(W1):W256–9.

105. SRA-Tools - NCBI [Internet]. [cited 2021 Aug 18]. Available from: http://ncbi.github.io/sra-tools/

106. Babraham Bioinformatics - Trim Galore! [Internet]. [cited 2021 Aug 18]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

107. Wickham H. ggplot2. WIREs Comput Stat. 2011;3(2):180–5.

108. Hullo MF, Auger S, Soutourina O, Barzu O, Yvon M, Danchin A, et al. Conversion of methionine to cysteine in Bacillus subtilis and its regulation. J Bacteriol. 2007 Jan;189(1):187–97.

109. Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S, et al. Expression Atlas update: from tissues to single cells. Nucleic Acids Res. 2020 Jan 8;48(D1):D77–83.

110. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020 Mar 1;36(6):1925–7.

111. Levy CC, Goldman P. The Enzymatic Hydrolysis of Methotrexate and Folic Acid. J Biol Chem. 1967 Jun 25;242(12):2933–8.

112. Ruiz JFM, Kedziora K, Windle H, Kelleher DP, Gilmer JF. Investigation into drug release from colon-specific azoreductase-activated steroid prodrugs using in-vitro models. J Pharm Pharmacol. 2011 Jun 1;63(6):806–16.

113. Marquez Ruiz JF, Kedziora K, Keogh B, Maguire J, Reilly M, Windle H, et al. A double prodrug system for colon targeting of benzenesulfonamide COX-2 inhibitors. Bioorg Med Chem Lett. 2011 Nov;21(22):6636–40.

114. Stenson WF, Lobos E. Sulfasalazine inhibits the synthesis of chemotactic lipids by neutrophils. J Clin Invest. 1982 Feb 1;69(2):494–7.

115. Weber CK, Liptay S, Wirth T, Adler G, Schmid RM. Suppression of NF-κB activity by sulfasalazine is mediated by direct inhibition of IκB kinases α and β. Gastroenterology. 2000 Nov 1;119(5):1209–18.

116. Rachmilewitz D, Karmeli F, Schwartz LW, Simon PL. Effect of aminophenols (5-ASA and 4-ASA) on colonic interleukin-1 generation. Gut. 1992 Jul 1;33(7):929–32.

117. Peppercorn MA, Goldman P. Distribution Studies of Salicylazosulfapyridine and its Metabolites. Gastroenterology. 1973 Feb 1;64(2):240–5.

118. Suzuki H. Remarkable diversification of bacterial azoreductases: primary sequences, structures, substrates, physiological roles, and biotechnological applications. Appl Microbiol Biotechnol. 2019 May 1;103(10):3965–78.

119. Brown JP. Reduction of Polymeric Azo and Nitro Dyes by Intestinal Bacteria. Appl Environ Microbiol. 1981 May 1;41(5):1283–6.

120. Rafii F, Cerniglia CE. Comparison of the azoreductase and nitroreductase from Clostridium perfringens. Appl Environ Microbiol. 1993 Jun;59(6):1731–4.

121. Rafii F, Cerniglia CE. Reduction of azo dyes and nitroaromatic compounds by bacterial enzymes from the human intestinal tract. Environ Health Perspect. 1995 Jun 1;103(suppl 5):17–9.

122. Liu G, Zhou J, Lv H, Xiang X, Wang J, Zhou M, et al. Azoreductase from Rhodobacter sphaeroides AS1.1737 is a flavodoxin that also functions as nitroreductase and flavin mononucleotide reductase. Appl Microbiol Biotechnol. 2007 Oct 1;76(6):1271–9.

123. Mercier C, Chalansonnet V, Orenga S, Gilbert C. Characteristics of major Escherichia coli reductases involved in aerobic nitro and azo reduction. J Appl Microbiol. 2013 Oct;115(4):1012–22.

124. Misal SA, Lingojwar DP, Lokhande MN, Lokhande PD, Gawai KR. Enzymatic transformation of nitro-aromatic compounds by a flavin-free NADH azoreductase from Lysinibacillus sphaericus. Biotechnol Lett. 2014 Jan 1;36(1):127–31.

125. Chalansonnet V, Mercier C, Orenga S, Gilbert C. Identification of Enterococcus faecalis enzymes with azoreductases and/or nitroreductase activity. BMC Microbiol. 2017 Dec;17(1):1–10.

126. Roldán MD, Pérez-Reinado E, Castillo F, Moreno-Vivián C. Reduction of polynitroaromatic compounds: the bacterial nitroreductases. FEMS Microbiol Rev. 2008 May 1;32(3):474–500.

127. Liu G, Zhou J, Jin R, Zhou M, Wang J, Lu H, et al. Enhancing survival of Escherichia coli by expression of azoreductase AZR possessing quinone reductase activity. Appl Microbiol Biotechnol. 2008 Sep 1;80(3):409.

128. Leelakriangsak M, Huyen NTT, Töwe S, Van Duy N, Becher D, Hecker M, et al. Regulation of quinone detoxification by the thiol stress sensing DUF24/MarR-like repressor, YodB in Bacillus subtilis. Mol Microbiol. 2008;67(5):1108–24.

129. Ryan A, Wang CJ, Laurieri N, Westwood I, Sim E. Reaction mechanism of azoreductases suggests convergent evolution with quinone oxidoreductases. Protein Cell. 2010 Aug 1;1(8):780–90.

130. Ryan A, Kaplan E, Nebel JC, Polycarpou E, Crescente V, Lowe E, et al. Identification of NAD(P)H Quinone Oxidoreductase Activity in Azoreductases from P. aeruginosa: Azoreductases and NAD(P)H Quinone Oxidoreductases Belong to the Same FMN-Dependent Superfamily of Enzymes. PLOS ONE. 2014 Jun 10;9(6):e98551.

131. Crescente V, Holland SM, Kashyap S, Polycarpou E, Sim E, Ryan A. Identification of novel members of the bacterial azoreductase family in Pseudomonas aeruginosa. Biochem J. 2016 Feb 24;473(5):549–58.

132. Ryan A. Azoreductases in drug metabolism. Br J Pharmacol. 2017 Jul;174(14):2161–73.

133. Stolz A. Basic and applied aspects in the microbial degradation of azo dyes. Appl Microbiol Biotechnol. 2001 Jul 1;56(1):69–80.

134. Chen H, Feng J, Kweon O, Xu H, Cerniglia CE. Identification and molecular characterization of a novel flavin-free NADPH preferred azoreductase encoded by azoB in Pigmentiphaga kullae K24. BMC Biochem. 2010 Mar 16;11(1):13.

135. Salem MA, Al-Ghonemiy AF, Zaki AB. Photocatalytic degradation of Allura red and Quinoline yellow with Polyaniline/TiO2 nanocomposite. Appl Catal B Environ. 2009 Sep 7;91(1):59–66.

136. Masilompane TM, Chaukura N, Mishra SB, Mishra AK. Chitosan-lignin-titania nanocomposites for the removal of brilliant black dye from aqueous solution. Int J Biol Macromol. 2018 Dec 1;120:1659–66.

137. Cerniglia CE, Freeman JP, Franklin W, Pack LD. Metabolism of azo dyes derived from benzidine, 3,3'-dimethylbenzidine and 3,3' -dimethoxybenzidine to potentially carcinogenic aromatic amines by intestinal bacteria. Carcinogenesis. 1982 Jan 1;3(11):1255–60.

138. Feng J, Cerniglia CE, Chen H. Toxicological significance of azo dye metabolism by human intestinal microbiota. Front Biosci Elite Ed. 2012 Jan 1;4:568–86.

139. Pandey A, Singh P, Iyengar L. Bacterial decolorization and degradation of azo dyes. Int Biodeterior Biodegrad. 2007 Mar 1;59(2):73–84.

140. Zimmermann T, Kulla HG, Leisinger T. Properties of Purified Orange II Azoreductase, the Enzyme Initiating Azo Dye Degradation by Pseudomonas KF46. Eur J Biochem. 1982;129(1):197–203.

141. Nakanishi M, Yatome C, Ishida N, Kitade Y. Putative ACP Phosphodiesterase Gene (acpD) Encodes an Azoreductase *. J Biol Chem. 2001 Dec 7;276(49):46394–9.

142. Suzuki Y, Yoda T, Ruhul A, Sugiura W. Molecular Cloning and Characterization of the Gene Coding for Azoreductase from Bacillus sp. OY1-2 Isolated from Soil *. J Biol Chem. 2001 Mar 23;276(12):9059–65.

143. Blümel S, Knackmuss HJ, Stolz A. Molecular Cloning and Characterization of the Gene Coding for the Aerobic Azoreductase from Xenophilus azovorans KF46F. Appl Environ Microbiol. 2002 Aug 1;68(8):3948–55.

144. Blümel S, Stolz A. Cloning and characterization of the gene coding for the aerobic azoreductase from Pigmentiphaga kullae K24. Appl Microbiol Biotechnol. 2003 Aug 1;62(2):186–90.

145. Chen H, Wang RF, Cerniglia CE. Molecular cloning, overexpression, purification, and characterization of an aerobic FMN-dependent azoreductase from Enterococcus faecalis. Protein Expr Purif. 2004 Apr 1;34(2):302–10.

146. Ooi T, Shibata T, Sato R, Ohno H, Kinoshita S, Thuoc TL, et al. An azoreductase, aerobic NADH-dependent flavoprotein discovered from Bacillus sp.: functional expression and enzymatic characterization. Appl Microbiol Biotechnol. 2007 May 1;75(2):377–86.

147. Matsumoto K, Mukai Y, Ogata D, Shozui F, Nduko JM, Taguchi S, et al. Characterization of thermostable FMN-dependent NADH azoreductase from the moderate thermophile Geobacillus stearothermophilus. Appl Microbiol Biotechnol. 2010 May 1;86(5):1431–8.

148. Misal SA, Lingojwar DP, Shinde RM, Gawai KR. Purification and characterization of azoreductase from alkaliphilic strain Bacillus badius. Process Biochem. 2011 Jun 1;46(6):1264–9.

149. Gonçalves AMD, Mendes S, de Sanctis D, Martins LO, Bento I. The crystal structure of Pseudomonas putida azoreductase – the active site revisited. FEBS J. 2013;280(24):6643–57.

150. Lang W, Sirisansaneeyakul S, Ngiwsara L, Mendes S, Martins LO, Okuyama M, et al. Characterization of a new oxygen-insensitive azoreductase from Brevibacillus laterosporus TISTR1911: Toward dye decolorization using a packed-bed metal affinity reactor. Bioresour Technol. 2013 Dec 1;150:298–306.

151. Zhang X, Ng IS, Chang JS. Cloning and characterization of a robust recombinant azoreductase from Shewanella xiamenensis BC01. J Taiwan Inst Chem Eng. 2016 Apr 1;61:97–105.

152. Eslami M, Amoozegar MA, Asad S. Isolation, cloning and characterization of an azoreductase from the halophilic bacterium Halomonas elongata. Int J Biol Macromol. 2016 Apr 1;85:111–6.

153. Sousa T, Yadav V, Zann V, Borde A, Abrahamsson B, Basit AW. On the colonic bacterial metabolism of azo-bonded prodrugsof 5-aminosalicylic acid. J Pharm Sci. 2014 Oct;103(10):3171–5.

154. Zahran SA, Ali-Tammam M, Ali AE, Aziz RK. Compositional variation of the human fecal microbiome in relation to azo-reducing activity: a pilot study. Gut Pathog. 2021 Oct 8;13(1):58.

155. Taffet SL, Das KM. Sulfasalazine. Dig Dis Sci. 1983 Sep 1;28(9):833–42.

156. Collins JR. Adverse reactions to salicylazosulfapyridine (azulfidine) in the treatment of ulcerative colitis. South Med J. 1968 Apr;61(4):354–8.

157. Das KM, Eastwood MA, McManus JPA, Sircus W. Adverse Reactions during Salicylazosulfapyridine Therapy and the Relation with Drug Metabolism and Acetylator Phenotype. N Engl J Med. 1973 Sep 6;289(10):491–5.

158. Taffet SL, Das KM. Desensitization of patients with inflammatory bowel disease to sulfasalazine. Am J Med. 1982 Oct;73(4):520–4.

159. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim Biophys Acta BBA - Proteins Proteomics. 2015 Aug 1;1854(8):1019–37.

160. Nagasawa T, Ishii T, Kumagai H, Yamada H. D-Cysteine desulfhydrase of Escherichia coli. Purification and characterization. Eur J Biochem. 1985 Dec 16;153(3):541–51.

161. Soutourina J, Blanquet S, Plateau P. Role of d-Cysteine Desulfhydrase in the Adaptation of Escherichia coli to d-Cysteine. J Biol Chem. 2001 Nov 2;276(44):40864–72.

162. Nonaka G, Takumi K. Cysteine degradation gene yhaM, encoding cysteine desulfidase, serves as a genetic engineering target to improve cysteine production in Escherichia coli. AMB Express. 2017 May 10;7(1):90.

163. Awano N, Wada M, Kohdoh A, Oikawa T, Takagi H, Nakamori S. Effect of cysteine desulfhydrase gene disruption on L-cysteine overproduction in Escherichia coli. Appl Microbiol Biotechnol. 2003 Aug;62(2–3):239–43.

164. Ng W, Tonzetich J. Effect of Hydrogen Sulfide and Methyl Mercaptan on the Permeability of Oral Mucosa. J Dent Res. 1984 Jul 1;63(7):994–7.

165. Wang J, Guo X, Li H, Qi H, Qian J, Yan S, et al. Hydrogen Sulfide From Cysteine Desulfurase, Not 3-Mercaptopyruvate Sulfurtransferase, Contributes to Sustaining Cell Growth and Bioenergetics in E. coli Under Anaerobic Conditions. Front Microbiol [Internet]. 2019 [cited 2021 Jan 19];10. Available from: http://www.frontiersin.org/articles/10.3389/fmicb.2019.02357/full

166. Andreeßen C, Gerlt V, Steinbüchel A. Conversion of cysteine to 3-mercaptopyruvic acid by bacterial aminotransferases. Enzyme Microb Technol. 2017 Apr 1;99:38–48.

167. Auger S, Gomez MP, Danchin A, Martin-Verstraete I. The PatB protein of Bacillus subtilis is a C-S-lyase. Biochimie. 2005 Feb 1;87(2):231–8.

168. Zdych E, Peist R, Reidl J, Boos W. MalY of Escherichia coli is an enzyme with the activity of a beta C-S lyase (cystathionase). J Bacteriol. 1995 Sep;177(17):5035–9.

169. Clausen T, Schlegel A, Peist R, Schneider E, Steegborn C, Chang YS, et al. X-ray structure of MalY from Escherichia coli: a pyridoxal 5′-phosphate-dependent enzyme acting as a modulator in mal gene expression. EMBO J. 2000 Mar 1;19(5):831–42.

170. Johnson PM, Beck CM, Morse RP, Garza-Sánchez F, Low DA, Hayes CS, et al. Unraveling the essential role of CysK in CDI toxin activation. Proc Natl Acad Sci U S A. 2016 Aug 30;113(35):9792–7.

171. DeMoss R. D., Moser K. Tryptophanase in Diverse Bacterial Species. J Bacteriol. 1969 Apr 1;98(1):167–71.

172. Schwartz CJ, Djaman O, Imlay JA, Kiley PJ. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. Proc Natl Acad Sci. 2000 Aug 1;97(16):9009–14.